Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

GEOMETRY

Notes Easter 2002

T. K. Carne.

t.k.carne@dpmms.cam.ac.uk

CONTENTS

0. INTRODUCTION

Geometry has been studied for a very long time. In Classical times the Greeks developed Geometry as their chief way to study Mathematics. About 300BC Euclid of Alexandria wrote the "Elements" that gathered together most of the Geometry that was known at that time and formulated it axiomatically. From the Renaissance until the 1950s, studying Euclid's Elements was an essential part of any Mathematical education. Unfortunately, it came to be taught in so pedestrian a way that people thought it was about memorising peculiar constructions rather than using pictures to guide mathematical thoughts. So it fell into contempt and now very little geometry is taught at school.

However, geometry has proved a vital part of many modern developments in Mathematics. It is clear that it is involved in studying the shapes of surfaces and topology. In physics, Einstein and Minkowski recognised that the laws of physics should be formulated in terms of the geometry of space-time. In algebra, one of the most fruitful ways to study groups is to represent them as symmetry groups of geometrical objects. In these cases, it is often not Euclidean geometry that is needed but rather hyperbolic geometry. Many of the courses in Part 2 use and develop geometrical ideas. This course is intended to give you a brief introduction to some of those ideas. We will use very little from earlier course beyond the Algebra and Geometry course in Part 1A. However, there will be frequent links to other courses. (Vector Calculus, Analysis, Further Analysis, Quadratic Mathematics, Complex Methods, Special Relativity.)

Euclid's 5th postulate for Geometry asserted that given any straight line ℓ in the Euclidean plane and any point P not on ℓ, there is an unique straight line through P that does not meet ℓ. This is the line through P parallel to ℓ. However, there are other surfaces than the Euclidean plane that have a different geometry. On the sphere, there are no two lines that do not intersect. In the hyperbolic plane, there are infinitely many lines through P that do not meet ℓ. In this course we will define the spherical (elliptic) and hyperbolic geometries and develop some of their simple properties.

Felix Klein (1849 – 1925) did a great deal of very beautiful Mathematics concerned with geometry. He saw that, in order to have an interesting geometry, it is crucial that there is a large group of symmetries acting on our space. We are then concerned with the geometrical properties of points and lines in the space; that is with properties that are invariant under the group of symmetries. For example, in Euclidean plane geometry we are concerned with those properties of configurations such as triangles that are unchanged when we apply rigid motions such as rotations and translations. This includes the lengths of the sides of the triangle and the angles at its vertices but not the particular co-ordinates of a vertex. This has the advantage that, when we wish to establish a geometrical property, we can move our triangle to any convenient location before beginning the proof. In this course, we will concentrate on groups of isometries that preserve a metric on the space.

The final, thirteenth, book of Euclid's Elements is about the five regular Platonic solids. This is both a beautiful piece of Mathematics and a very important example of many techniques in geometry. So we too will study them. We will prove that that there are only five, study their symmetry groups, and briefly consider the corresponding results in hyperbolic space.

The books recommended in the schedules are all good. My favourite is:

Elmer Rees, *Notes on Geometry*, Springer Universitext, 1998

which is suitably short. In addition, I also suggest

H.S.M. Coxeter, *Introduction to Geometry*, 2nd Edition, Wiley Classics, 1989.

This gives a gentle introduction to a broad vista of geometry and is written by one of the current masters of geometry.

1. EUCLIDEAN GEOMETRY

1.1 Euclidean Space

Euclidean geometry is concerned with the properties of points and lines in Euclidean N-space. In the Algebra and Geometry course you studied the vector space $\mathbb{R}^N$. We will use this as a model for Euclidean Geometry of *Euclidean N-space* $\mathbb{E}^N$.

The points of $\mathbb{E}^N$ are the elements of $\mathbb{R}^N$ and the (straight) lines of $\mathbb{E}^N$ are the subsets

$$\{\boldsymbol{x} \in \mathbb{R}^N : \boldsymbol{x} = \boldsymbol{A} + \lambda \boldsymbol{u} \text{ for some } \lambda \in \mathbb{R}\}$$

for a non-zero vector $\boldsymbol{u}$. Note that any non-zero multiple $\lambda \boldsymbol{u}$ gives the same line as $\boldsymbol{u}$. For any two distinct points $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^N$, there is an unique line through them: $\{\boldsymbol{A} + \lambda(\boldsymbol{B} - \boldsymbol{A}) : \lambda \in \mathbb{R}\}$. Two lines $\{\boldsymbol{A} + \lambda \boldsymbol{u} : \lambda \in \mathbb{R}\}$ and $\{\boldsymbol{B} + \mu \boldsymbol{v} : \mu \in \mathbb{R}\}$ are *parallel* if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent (that is, each is multiple of the other). In this case, the two lines are either disjoint or identical. (Since $\mathbb{R}^N$ and $\mathbb{E}^N$ have the same points, we could identify them. This is often done but you should note that the inner product space $\mathbb{R}^N$ has many more properties than $\mathbb{E}^N$. For example, in $\mathbb{R}^N$ we can form the sum of two points but in $\mathbb{E}^N$ the sum of two points is not well defined since it changes under isometries.)

The *inner product* (dot product) of two vectors in $\mathbb{R}^N$ is

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{n=1}^N x_n y_n \ .$$

The *norm* or *length* of a vector is

$$||\boldsymbol{x}|| = (\boldsymbol{x} \cdot \boldsymbol{x})^{1/2} = \left(\sum_{n=1}^N x_n^2\right)^{1/2} \ .$$

We define the distance between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{E}^N$ to be

$$d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x} - \boldsymbol{y}|| \ .$$

This gives a metric on $\mathbb{E}^N$ called the *Euclidean metric*.

A line can always be written as $\{\boldsymbol{A} + \lambda \boldsymbol{u} : \lambda \in \mathbb{R}\}$ for a unit vector $\boldsymbol{u} \in \mathbb{R}^N$. The unit vector $\boldsymbol{u}$ is uniquely determined up to a change of sign. To allow for this we can mark the line with an arrow pointing in the direction of $\boldsymbol{u}$. We call this a *directed line* and $\boldsymbol{u}$ is the unique unit vector giving the direction of the line. Let ℓ and m be two directed lines with direction vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively. The *angle* between the two directed lines is the unique number $\theta \in [0, \pi]$ with

$$\boldsymbol{u} \cdot \boldsymbol{v} = \cos\theta \ .$$

Lines are parallel when the angle between them is 0 or π.

Consider two different points $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\mathbb{E}^N$. A *path* from $\boldsymbol{A}$ to $\boldsymbol{B}$ is a smooth map

$$\gamma : [0, 1] \to \mathbb{E}^N \ ; \quad t \mapsto \gamma(t)$$

with $\gamma(0) = \boldsymbol{A}$ and $\gamma(1) = \boldsymbol{B}$. Its *length* is

$$L(\gamma) = \int_0^1 ||\gamma'(t)|| \, dt \ .$$

For each $t \in [0, 1]$, the derivative $\gamma'(t)$ is a vector in $\mathbb{R}^N$ that is tangent to the path at $\gamma(t)$. Since
$$||\gamma(1) - \gamma(0)|| = \left|\left| \int_0^1 \gamma'(t)\, dt \right|\right| \leqslant \int_0^1 ||\gamma'(t)||\, dt = L(\gamma)$$
we see that the shortest length of a path from $\boldsymbol{A}$ to $\boldsymbol{B}$ is the distance $d(\boldsymbol{A}, \boldsymbol{B})$ and this length occurs only when γ traces out the straight line segment
$$[\boldsymbol{A}, \boldsymbol{B}] = \{(1 - t)\boldsymbol{A} + t\boldsymbol{B} : t \in [0, 1]\} .$$

If $\boldsymbol{u}$ is the unit vector pointing from $\boldsymbol{A}$ towards $\boldsymbol{B}$, then the path
$$\gamma : t \mapsto \boldsymbol{A} + t\boldsymbol{u}$$
traces out the line through $\boldsymbol{A}$ and $\boldsymbol{B}$ as t runs through $\mathbb{R}$. As t runs from 0 to $d(\boldsymbol{A}, \boldsymbol{B})$, so γ traces out the shortest path from $\boldsymbol{A}$ to $\boldsymbol{B}$. We call γ the *geodesic* from $\boldsymbol{A}$ to $\boldsymbol{B}$ in the Euclidean N-space. Thus we can think of a straight line as giving the shortest path or geodesic between any two points on it.

1.2 Euclidean Isometries

An *isometry of* $\mathbb{E}^N$ is a map $T : \mathbb{E}^N \to \mathbb{E}^N$ that preserves the Euclidean metric:
$$d(T(\boldsymbol{x}), T(\boldsymbol{y})) = d(\boldsymbol{x}, \boldsymbol{y}) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{E}^N .$$
For example, *translation* by a vector $\boldsymbol{t} \in \mathbb{R}^N$: $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{t}$ is an isometry. Also, if R is an orthogonal $N \times N$ matrix (so $R^t R = I$), then the map $\boldsymbol{x} \mapsto R\boldsymbol{x}$ is also an isometry. The next result shows that the only isometries are the ones obtained by combining these two.

The set $\text{Isom}(\mathbb{E}^N)$ of all isometries of $\mathbb{E}^N$ forms a group. We will be interested in properties that are invariant under the elements of this group.

Proposition 1.1 Isometries of $\mathbb{E}^N$
Every isometry of the Euclidean N-space $\mathbb{E}^N$ is of the form
$$\boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$$
for some $R \in \mathrm{O}(N)$ and $\boldsymbol{t} \in \mathbb{R}^N$. Moreover, every such map is an isometry of $\mathbb{E}^N$.

Proof:
    It is clear that every such map preserves the inner product and hence the Euclidean metric.

Suppose now that $T : \mathbb{E}^N \to \mathbb{E}^N$ is an isometry. Set $\boldsymbol{t} = T(\boldsymbol{0})$ and consider
$$S : \mathbb{E}^N \to \mathbb{E}^N ; \quad \boldsymbol{x} \mapsto T(\boldsymbol{x}) - \boldsymbol{t} .$$
Since T is an isometry, so is S and we also have $S(\boldsymbol{0}) = \boldsymbol{0}$.

For two vectors $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^N$ we have
$$||\boldsymbol{A} - \boldsymbol{B}||^2 = ||\boldsymbol{A}||^2 + ||\boldsymbol{B}||^2 - 2\boldsymbol{A} \cdot \boldsymbol{B} .$$
So $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthonormal if and only if
$$||\boldsymbol{A}|| = ||\boldsymbol{B}|| = 1 \quad \text{and} \quad ||\boldsymbol{A} - \boldsymbol{B}|| = \sqrt{2} .$$
Now let $\boldsymbol{f}_n = S(\boldsymbol{e}_n)$ be the image of the standard basis vectors $\boldsymbol{e}_n$ of $\mathbb{R}^N$ under S. Then the above argument shows that $(\boldsymbol{f}_n)$ is an orthonormal basis for $\mathbb{R}^N$. Consequently, there is an orthogonal matrix R with $R\boldsymbol{e}_n = \boldsymbol{f}_n$.

Finally, consider $\boldsymbol{x} = \sum x_n \boldsymbol{e}_n \in \mathbb{E}^N$. This satisfies
$$||\boldsymbol{x} - \boldsymbol{e}_n||^2 - ||\boldsymbol{x}||^2 - ||\boldsymbol{e}_n||^2 = 2\boldsymbol{x} \cdot \boldsymbol{e}_n = 2x_n .$$
Since S is an isometry, we have
$$||S(\boldsymbol{x}) - \boldsymbol{f}_n||^2 - ||S(\boldsymbol{x})||^2 - ||\boldsymbol{f}_n||^2 = 2x_n$$
and therefore $S(\boldsymbol{x}) = \sum x_n \boldsymbol{f}_n = \sum x_n S(\boldsymbol{e}_n)$. Thus S is linear.

Since S and R agree on the standard basis, they must be identical. Therefore,
$$T(\boldsymbol{x}) = S(\boldsymbol{x}) + \boldsymbol{t} = R\boldsymbol{x} + \boldsymbol{t}$$
as required. $\qquad\qquad \square$

Note that this proposition certainly shows that any isometry maps a straight line onto another straight line, and that it preserves the angle between two lines.

The orthogonal matrices $R \in \mathrm{O}(N)$ have determinant $+1$ or -1. Those with determinant $+1$ preserve the orientation of a basis while those with determinant -1 reverse the orientation. We write $\mathrm{SO}(N)$ for the set of matrices in $\mathrm{O}(N)$ with determinant $+1$. This is the *special orthogonal group*. The map

$$\varepsilon : \mathrm{O}(N) \to \{-1, +1\} \; ; \quad R \mapsto \det R$$

is a group homomorphism so its kernel, $\mathrm{SO}(N)$, is a normal subgroup of $\mathrm{O}(N)$ with index 2. In a similar way, we will say that an isometry $T \in \mathrm{Isom}(\mathbb{E}^N)$ is *orientation preserving* if $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ with $R \in \mathrm{SO}(N)$. These form a subgroup $\mathrm{Isom}^+(\mathbb{E}^N)$ of $\mathrm{Isom}(\mathbb{E}^N)$. If $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ with $R \in \mathrm{O}(N)\backslash\mathrm{SO}(N)$, then we say that T is *orientation reversing*. The map

$$\varepsilon : \mathrm{Isom}(\mathbb{E}^N) \to \{-1, +1\}$$

that sends the isometry $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ to $\det R$ is a group homomorphism, so its kernel, $\mathrm{Isom}^+(\mathbb{E}^N)$, is a normal subgroup of $\mathrm{Isom}(\mathbb{E}^N)$ of index 2.

In low dimensions we can be more specific about the nature of isometries.

Proposition 1.2   Isometries of $\mathbb{E}^2$
An orientation preserving isometry of the Euclidean plane $\mathbb{E}^2$ is:

(a) *The identity.*

(b) *A translation.*

(c) *A rotation about some point $\boldsymbol{c} \in \mathbb{E}^2$.*

An orientation reversing isometry of $\mathbb{E}^2$ is:

(d) *A reflection.*

(e) *A glide reflections, that is a reflection in a line ℓ followed by a translation parallel to ℓ.*

Proof:
Consider first the isometries of $\mathbb{E}^2$ that fix the origin. These are the orthogonal maps $\boldsymbol{x} \mapsto R\boldsymbol{x}$ for $R \in \mathrm{O}(2)$. The orientation preserving orthogonal linear maps are:

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \text{for} \quad \theta \in [0, 2\pi) \; .$$

These are the rotations about the origin through an angle θ. The eigenvalues of R are $e^{i\theta}$ and $e^{-i\theta}$. Similarly, the orientation reversing orthogonal linear maps are:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \qquad \text{for} \quad \theta \in [0, 2\pi) \; .$$

These are reflections in the line $\{\lambda(\cos\frac{1}{2}\theta, \sin\frac{1}{2}\theta) : \lambda \in \mathbb{R}\}$ that passes through the origin. The eigenvalues of R are 1 and -1.

An isometry T of $\mathbb{E}^2$ is of the form $\boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ with $R \in \mathrm{O}(2)$ and $\boldsymbol{t} \in \mathbb{R}^2$. Suppose that T is orientation preserving. Then $R \in \mathrm{SO}(2)$ so either R is the identity I, or R is a rotation about the origin

through some angle θ. If $R = I$ then we obtain case (a) when $\boldsymbol{t} = \boldsymbol{0}$ and case (b) otherwise. If R is a rotation, then 1 is not an eigenvalue of R, so there is a vector $\boldsymbol{c}$ with $(I - R)\boldsymbol{c} = \boldsymbol{t}$. Hence,

$$T(\boldsymbol{x}) = R\boldsymbol{x} + \boldsymbol{t} = R(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{c}$$

and T is a rotation through an angle θ about the centre $\boldsymbol{c}$. This is case (c).

Now suppose that T is orientation reversing, so $\det R = -1$. Then R is a reflection in some line through the origin, say m. The linear map $I - R$ maps onto the 1-dimensional vector subspace of $\mathbb{R}^2$ orthogonal to m, so we can find a vector $\boldsymbol{c} \in \mathbb{R}^2$ with $\boldsymbol{u} = \boldsymbol{t} - (I - R)\boldsymbol{c}$ parallel to m. Hence,

$$T(\boldsymbol{x}) = R\boldsymbol{x} + \boldsymbol{t} = (R(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{c}) + \boldsymbol{u} .$$

If $\boldsymbol{u} = \boldsymbol{0}$, then T is a reflection in the line ℓ obtained by translating m by $\boldsymbol{c}$. If $\boldsymbol{u} \neq \boldsymbol{0}$, then T is the glide reflection obtained by following reflection in ℓ by a translation of $\boldsymbol{u}$ parallel to ℓ. $\qquad \square$

We can analyse isometries of $\mathbb{E}^3$ in a similar way. First we need to describe some examples. The map

$$R : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

is a rotation through an angle θ about the x-axis (or the identity if θ is an integer multiple of 2π). For any (orientation preserving) isometry T of $\mathbb{E}^3$ we call the conjugate TRT^{-1} a rotation through an angle θ about the line m that is the image of the x-axis under T. Such a map fixes each point of m and acts on each plane orthogonal to m as a rotation through an angle θ. Similarly, the linear map

$$R : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

is obtained by rotating through an angle θ about the x-axis and then reflecting in the plane perpendicular to the x-axis. For any (orientation preserving) isometry T of $\mathbb{E}^3$ the conjugate TRT^{-1} is a rotation through an angle θ about the line m that is the image of the x-axis under T followed by reflection in the plane through $T(\boldsymbol{0})$ perpendicular to m. If θ is an integer multiple of 2π then this map is just a reflection in the plane. Otherwise, we call it a *rotatory reflection* about the axis m.

These examples give us all of the orthogonal linear maps of $\mathbb{R}^3$:

Lemma 1.3 Orthogonal linear maps in $\mathbb{R}^3$
Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be an orthogonal linear map. If R is orientation preserving, then either R is the identity or else a rotation about a line m through the origin. If R is orientation reversing, then R is either a reflection in a plane through the origin or else a rotation about an line through the origin followed by reflection in the plane through the origin perpendicular to that line.

Proof:
We know that the matrix R has three complex eigenvalues, possibly coincident, each of modulus 1. Since the characteristic polynomial of R has real coefficients, these must occur in complex conjugate pairs. So at least one must be either $+1$ or -1. The product of the complex conjugate pairs is 1 and the product of all the eigenvalues is $\det R$, so there must be an eigenvalue $\varepsilon = \det R = \pm 1$.

Let $\boldsymbol{f}_1$ be a unit eigenvector with eigenvalue ε. We can extend $\boldsymbol{f}_1$ to an orthonormal basis $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3$ for $\mathbb{R}^3$. Then R is of the form

$$R : \lambda_1 \boldsymbol{f}_1 + \lambda_2 \boldsymbol{f}_2 + \lambda_3 \boldsymbol{f}_3 \mapsto \varepsilon \lambda_1 \boldsymbol{f}_1 + (a\lambda_2 + b\lambda_3)\boldsymbol{f}_2 + (c\lambda_2 + d\lambda_3)\boldsymbol{f}_3$$

for some real numbers a, b, c, d. Since R is orthogonal, the matrix $R_o = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ must also be orthogonal. Also, $\det R = \varepsilon \det R_o$ so R_o has determinant $+1$. Thus Proposition 1.2 shows us that R_o is either the

identity or a rotation of a plane about the origin through an angle θ. It follows that R itself is one of the following:

(i) the identity,

(ii) a rotation about $\boldsymbol{f}_1$ through an angle θ,

when R is orientation preserving and one of

(iii) a reflection in the plane spanned by $\boldsymbol{f}_2, \boldsymbol{f}_3$,

(iv) a rotation about $\boldsymbol{f}_1$ through an angle θ followed by a reflection in the plane spanned by $\boldsymbol{f}_2, \boldsymbol{f}_3$,

when R is orientation reversing. $\qquad\square$

Proposition 1.4   Isometries of $\mathbb{E}^3$
An orientation preserving isometry of Euclidean 3-space $\mathbb{E}^3$ is:

(a) *The identity.*

(b) *A translation.*

(c) *A rotation about some line ℓ.*

(d) *A screw rotation, that is a rotation about some line ℓ followed by a translation parallel to ℓ.*

An orientation reversing isometry of $\mathbb{E}^3$ is:

(e) *A reflection in some plane Π.*

(f) *A glide reflection, that is a reflection in a plane Π followed by a translation parallel to Π.*

(g) *A rotatory reflection, that is a rotation about some axis ℓ followed by reflection in a plane perpendicular to ℓ.*

Proof:

Let $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ be the isometry of $\mathbb{E}^3$. The Lemma tells us the possibilities for R.

Consider first the case where T is orientation preserving. Then $\det R = +1$ and R is either the identity or a rotation about a line m. If $R = I$, then we have case (a) when $\boldsymbol{t} = \boldsymbol{0}$ and case (b) if $\boldsymbol{t} \neq \boldsymbol{0}$. Otherwise, R is a rotation about m and so $I - R$ maps onto the plane $m^\perp$ perpendicular to m. Hence, we can find $\boldsymbol{c} \in \mathbb{R}^3$ with $\boldsymbol{u} = \boldsymbol{t} - (I - R)\boldsymbol{c}$ parallel to m. Then

$$T(\boldsymbol{x}) = (R(\boldsymbol{x} + \boldsymbol{c}) - \boldsymbol{c}) + \boldsymbol{u} \ .$$

So T is a rotation about $m + \boldsymbol{c}$ when $\boldsymbol{u} = \boldsymbol{0}$ and a screw rotation about $m + \boldsymbol{c}$ otherwise.

Now suppose that $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ is an orientation reversing isometry of $\mathbb{E}^3$. Then $\det R = -1$ and the Lemma gives the possibilities for R. There is a line m through the origin and a plane $m^\perp$ perpendicular to m with either R reflection in $m^\perp$ or R a rotation about m followed by reflection in $m^\perp$. If R is a reflection, then $I - R$ maps onto m, so we can find $\boldsymbol{c}$ with $\boldsymbol{u} = (I - R)\boldsymbol{c} + \boldsymbol{t}$ parallel to m. Then

$$T(\boldsymbol{x}) = (R(\boldsymbol{x} + \boldsymbol{c}) - \boldsymbol{c}) + \boldsymbol{u} \ .$$

If $\boldsymbol{u} = \boldsymbol{0}$, then T is reflection in $m^\perp - \boldsymbol{c}$. If $\boldsymbol{u} \neq \boldsymbol{0}$, then T is a glide reflection.

If R is reflection in $m^\perp$ followed by a rotation about m, then $I - R$ maps onto all of $\mathbb{R}^3$. So we can find $\boldsymbol{c}$ with $(I - R)\boldsymbol{c} + \boldsymbol{t} = \boldsymbol{0}$. Then

$$T(\boldsymbol{x}) = R(\boldsymbol{x} + \boldsymbol{c}) - \boldsymbol{c} \ .$$

So T is reflection in the plane $m^\perp - \boldsymbol{c}$ followed by a rotation about $m - \boldsymbol{c}$. $\qquad\square$

1.3 Euclidean Triangles

In this section we will study the geometry of the Euclidean plane $\mathbb{E}^2$. Let $\ell = \{\boldsymbol{A} + \lambda \boldsymbol{u} : \lambda \in \mathbb{R}\}$ be a line in $\mathbb{E}^2$ and $\boldsymbol{P}$ a point not lying on ℓ. Then there is an unique line through $\boldsymbol{P}$ parallel to ℓ given by $\{\boldsymbol{P} + \lambda \boldsymbol{u} : \lambda \in \mathbb{R}\}$. For any other line through $\boldsymbol{P}$, say $m = \{\boldsymbol{P} + \mu \boldsymbol{v} : \mu \in \mathbb{R}\}$ there is an unique point where ℓ and m meet. For $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent, so they span $\mathbb{R}^2$ and we can find $\lambda, \mu \in \mathbb{R}$ with $\boldsymbol{P} - \boldsymbol{A} = \lambda \boldsymbol{u} - \mu \boldsymbol{v}$.

Let $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ be three distinct points in $\mathbb{E}^2$. These form the vertices of a triangle Δ with the line segments $[\boldsymbol{B}, \boldsymbol{C}]$, $[\boldsymbol{C}, \boldsymbol{A}]$ and $[\boldsymbol{A}, \boldsymbol{B}]$ as the sides. We will denote the lengths of these sides by a, b and c respectively. The angles of the triangle at $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ will be denoted by α, β and γ respectively. So α is the angle between the vectors from $\boldsymbol{A}$ to $\boldsymbol{B}$ and from $\boldsymbol{A}$ to $\boldsymbol{C}$.
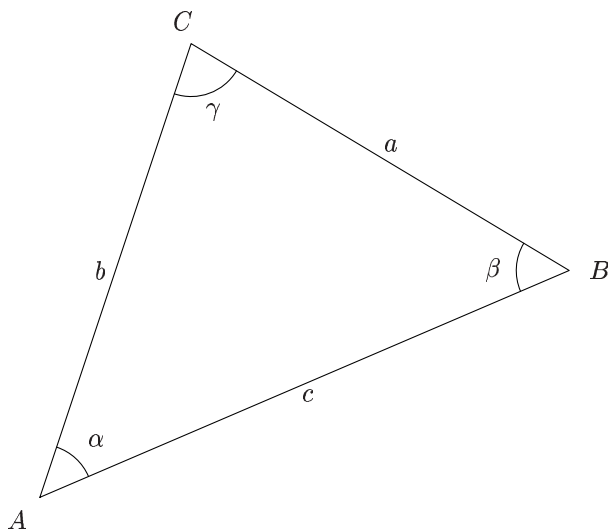
We say that two triangles are isometric if there is an isometry $T : \mathbb{E}^2 \to \mathbb{E}^2$ that maps one onto the other.

Proposition 1.5   Side lengths determine an Euclidean triangle up to isometry
Two triangles Δ, Δ' in the Euclidean plane $\mathbb{E}^2$ are isometric if and only if they have the same side lengths.

Proof:
Let Δ have vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and Δ' have vertices $\boldsymbol{A}', \boldsymbol{B}', \boldsymbol{C}'$. If T is an isometry with $T(\boldsymbol{A}) = \boldsymbol{A}'$, $T(\boldsymbol{B}) = \boldsymbol{B}'$ and $T(\boldsymbol{C}) = \boldsymbol{C}'$ then the side lengths of Δ' are

$$a' = d(\boldsymbol{B}', \boldsymbol{C}') = d(T(\boldsymbol{B}), T(\boldsymbol{C})) = d(\boldsymbol{B}, \boldsymbol{C}) = a \; , \quad etc.$$

So the side lengths of Δ' are the same as those of Δ.

Conversely, suppose that Δ' has the same side lengths as Δ. Then there is a translation T_1 that sends $\boldsymbol{A}$ to $\boldsymbol{A}'$. The points $T_1(\boldsymbol{B})$ and $\boldsymbol{B}'$ are both at distance c from $\boldsymbol{A}'$, so there is a rotation T_2 about $\boldsymbol{A}'$ that maps $T_1(\boldsymbol{B})$ onto $\boldsymbol{B}'$. Hence the isometry $S = T_2 T_1$ maps $\boldsymbol{A}$ to $\boldsymbol{A}'$ and $\boldsymbol{B}$ to $\boldsymbol{B}'$. Both the point $S(\boldsymbol{C})$ and $\boldsymbol{C}'$ are at distance a from $\boldsymbol{B}'$ and distance b from $\boldsymbol{C}'$. There are at most two points with this property and they are reflections of one another in the line through $\boldsymbol{B}'$ and $\boldsymbol{C}'$. Hence, either $\boldsymbol{C}' = S(\boldsymbol{C})$ or $\boldsymbol{C}' = RS(\boldsymbol{C})$ where R is the reflection in the line through $\boldsymbol{B}'$ and $\boldsymbol{C}'$. Thus, there is an isometry T, equal to either S or RS, with $T(\boldsymbol{A}) = \boldsymbol{A}'$, $T(\boldsymbol{B}) = \boldsymbol{B}'$ and $T(\boldsymbol{C}) = \boldsymbol{C}'$. This implies that $T(\Delta) = \Delta'$. $\qquad\square$

Two isometric triangles also have the same angles but the converse fails. Two triangles Δ and Δ' that have the same triangles are *similar*, that is, there is an enlargement of Δ that is isometric to Δ'.

When we follow the sides of a triangle in order around that triangle, we turn through one complete revolution. This shows that:

Proposition 1.6   Sum of angles of an Euclidean triangle
The sum of the angles of an Euclidean triangle is π.

Proof:
　　　　Let the triangle be Δ with vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and angles α, β, γ. By reflecting the triangle, if necessary, we may ensure that the vertices $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ occur in order as we go anti-clockwise around the triangle. Then

the angle from $\boldsymbol{AB}$ to $\boldsymbol{BC}$ is $\pi - \beta$,
the angle from $\boldsymbol{BC}$ to $\boldsymbol{CA}$ is $\pi - \gamma$,
the angle from $\boldsymbol{CA}$ to $\boldsymbol{AB}$ is $\pi - \alpha$,

and these angles are all measured in the positive (anti-clockwise) direction. Summing these shows that $(\pi - \alpha) + (\pi - \beta) + (\pi - \gamma) = 2\pi$. □

Since the side lengths a, b, c of a triangle Δ determine it up to isometry, they must also determine the angles. The cosine rule allows us to calculate these angles in terms of the side lengths.

Proposition 1.7   Euclidean Cosine rule
For an Euclidean triangle Δ
$$a^2 = b^2 + c^2 - 2bc\cos\alpha \ .$$

Proof:
　　　　We will prove this by using Cartesian co-ordinates in $\mathbb{R}^2$. To simplify matters, we first use an isometry to move Δ so the vertices are easier to work with. We can apply an isometry to Δ to move $\boldsymbol{A}$ to the origin. Hence we may assume that $\boldsymbol{A} = \boldsymbol{0}$. Now $\boldsymbol{B}$ is at distance c from $\boldsymbol{0}$, so we may rotate about $\boldsymbol{0}$ until $\boldsymbol{B} = (c, 0)$. Finally, $\boldsymbol{C}$ is at distance b from $\boldsymbol{0}$ and the angle at $\boldsymbol{A} = \boldsymbol{0}$ is α. Therefore, either $\boldsymbol{C} = (b\cos\alpha, b\sin\alpha)$ or $\boldsymbol{C} = (b\cos\alpha, -b\sin\alpha)$. In the latter case we reflect Δ in the x-axis. After doing this we have a triangle Δ, isometric to the original one, with

$$\boldsymbol{A} = \boldsymbol{0} \ , \quad \boldsymbol{B} = (c, 0) \ , \quad \boldsymbol{C} = (b\cos\alpha, b\sin\alpha) \ .$$

Now a simple calculation gives:

$$a^2 = d(\boldsymbol{B}, \boldsymbol{C})^2 = (b\cos\alpha - c)^2 + (b\sin\alpha)^2 = b^2 + c^2 - 2bc\cos\alpha \ .$$

□

Note that we get other forms of the cosine rule by permuting the vertices of Δ. The case where Δ has a right-angle at $\boldsymbol{A}$ gives Pythagoras' theorem: $a^2 = b^2 + c^2$.

Proposition 1.8   The Euclidean Sine rule
For an Euclidean triangle Δ
$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma} \ .$$

Proof:

Let $\boldsymbol{C}^*$ be the reflection of $\boldsymbol{C}$ in the line through $\boldsymbol{A}$ and $\boldsymbol{B}$. As in the proof of the cosine rule, we may assume that

$$\boldsymbol{A} = \boldsymbol{0} \ , \quad \boldsymbol{B} = (c, 0) \ , \quad \boldsymbol{C} = (b\cos\alpha, b\sin\alpha) \quad \text{and then} \quad \boldsymbol{C}' = (b\cos\alpha, -b\sin\alpha)$$

in order to calculate $d(\boldsymbol{C}, \boldsymbol{C}^*)$. This shows that $d(\boldsymbol{C}, \boldsymbol{C}^*) = 2b\sin\alpha$.

If we interchange $\boldsymbol{A}$ and $\boldsymbol{B}$ the point $\boldsymbol{C}^*$ is unchanged, so we must have

$$2b\sin\alpha = 2a\sin\beta \ .$$

Therefore,

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} \ .$$

$\square$

2. THE SPHERE

2.1 The geometry of the sphere

We will study the geometry of the unit sphere $S^2 = \{x \in \mathbb{R}^3 : ||x|| = 1\}$. A plane in $\mathbb{R}^3$ through the origin cuts S^2 in a circle of radius 1. This is called a *great circle* and plays the role of a straight line in Euclidean geometry. So we will call a great circle a *spherical line*. Two position vectors $P, Q \in S^2$ span a plane unless $P = \pm Q$. Hence there is always a spherical line through P and Q and the line is unique unless $P = Q$ or $P = -Q$. In the latter case we say that P and Q are *antipodal*. Such a spherical line consists of two arcs from P to Q. The shorter of these two arcs has length $\cos^{-1} P \cdot Q$. We will see that this arc is the shortest path from P to Q on the sphere and that $d(\cdot, \cdot)$ gives a metric on the sphere.

Any (smooth) path $\gamma : [a, b] \rightarrow S^2$ is also a path in $\mathbb{R}^3$ and so has a length $L(\gamma)$. We will prove that there is a shortest path from P to Q on S^2. First note that the length of a path is unchanged by rotations of $\mathbb{R}^3$ about axes through the origin. By applying such a rotation we may ensure that $\gamma(a) = P = (0, 0, 1)$ and $\gamma(b) = Q = (\sin \alpha, 0, \cos \alpha)$ with $\alpha = d(P, Q) = \cos^{-1} P \cdot Q$. We can write γ in cylindrical polar co-ordinates as

$$\gamma(t) = (r(t) \cos \theta(t), r(t) \sin \theta(t), z(t))$$

(where $r(t)^2 + z(t)^2 = 1$). So

$$||\gamma'(t)||^2 = r'(t)^2 + r(t)^2 \theta'(t)^2 + z'(t)^2 .$$

Consider the new curve $\beta(t) = (r(t), 0, z(t))$. This also goes from P to Q but lies on a spherical line. Furthermore

$$||\beta'(t)||^2 = r'(t)^2 + z'(t)^2 .$$

So $||\beta'(t)|| \leqslant ||\gamma'(t)||$ and $L(\beta) \leqslant L(\gamma)$. Indeed, the path β is shorter than γ unless $||\beta'(t)|| = ||\gamma'(t)||$ for all $t \in [a, b]$. This occurs only when θ is constant. It is now easy to see that the shortest path from $(0, 0, 1)$ to $(\sin \alpha, 0, \cos \alpha)$ is one that goes monotonically along the arc of the great circle

$$[0, \alpha] \rightarrow S^2 ; \quad t \mapsto (\cos t, 0, \sin t) .$$

This has length α.

Given any path of shortest length from P to Q in S^2, we can re-parametrize it by arc length. We call the resulting path a *geodesic* from P to Q in S^2. We have shown that each geodesic from P to Q is of the form

$$\gamma : [0, \alpha] \rightarrow S^2 ; \quad t \mapsto (\cos t) P + (\sin t) u$$

where u is a unit vector perpendicular to P. The vector u is the tangent vector to the geodesic at P. This geodesic is unique except when P and Q are antipodal. The distance $d(P, Q)$ is the length α of this geodesic.
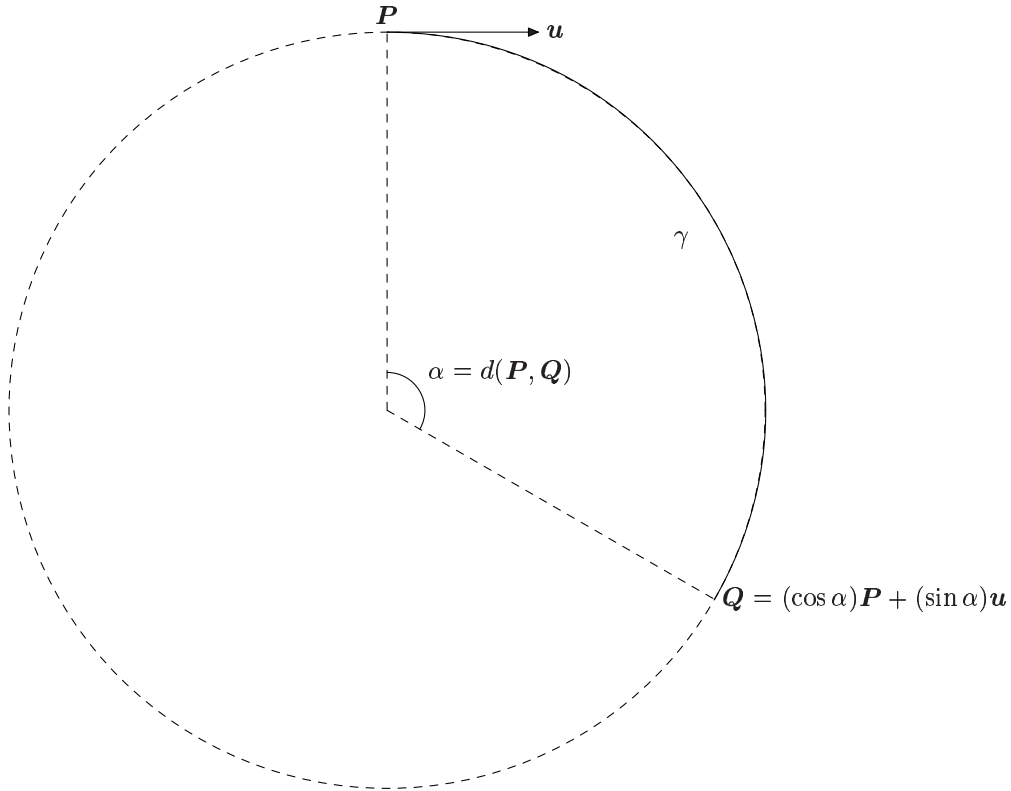
Proposition 2.1   Geodesics on the sphere
For any two points $P, Q \in S^2$, the shortest path from P to Q follows the shorter arc of a spherical line through P and Q. The length of this path is

$$d(P, Q) = \cos^{-1} P \cdot Q .$$

This gives a metric $d(\cdot, \cdot)$ on the sphere.

Proof:
        We have already proved the formula for $d(P, Q)$. It remains to show that d is a metric. It is clear that $d(P, Q) \geqslant 0$, with equality if and only if $P = Q$. Also $d(P, Q) = d(Q, P)$. Finally, if R is another point on S^2, then following the geodesic from P to Q and then the geodesic from Q to R gives a path of length $d(P, Q) + d(Q, R)$ from P to R. So we have the triangle inequality:

$$d(P, Q) + d(Q, R) \geqslant d(P, R)$$

as required.   □

        We will always use this metric d on the sphere.

2.2 Spherical isometries

Proposition 2.2   Isometries of S^2
Every isometry of S^2 is of the form $x \mapsto R(x)$ for an orthogonal linear map $R : \mathbb{R}^3 \to \mathbb{R}^3$.

Proof:

It is clear that every orthogonal linear map $R : \mathbb{R}^3 \to \mathbb{R}^3$ maps the unit sphere onto itself and preserves the metric $d(\boldsymbol{P}, \boldsymbol{Q}) = \cos^{-1} \boldsymbol{P} \cdot \boldsymbol{Q}$.

Suppose that $T : S^2 \to S^2$ is an isometry of the sphere. Proposition 2.1 shows that
$$T(\boldsymbol{P}) \cdot T(\boldsymbol{Q}) = \cos d(T(\boldsymbol{P}), T(\boldsymbol{Q})) = \cos d(\boldsymbol{P}, \boldsymbol{Q}) = \boldsymbol{P} \cdot \boldsymbol{Q} \ .$$
So T preserves the inner product of vectors in S^2. Let $\boldsymbol{e}_1, \boldsymbol{e_2}, \boldsymbol{e}_3$ be an orthonormal basis for $\mathbb{R}^3$. Since T preserves inner products, $T(\boldsymbol{e}_1), T(\boldsymbol{e}_2), T(\boldsymbol{e}_3)$ is also an orthonormal basis. Hence there is an orthogonal linear map $R : \mathbb{R}^3 \to \mathbb{R}^3$ with $T(\boldsymbol{e}_j) = R(\boldsymbol{e}_j)$ for $j = 1, 2, 3$. For any vector $\boldsymbol{x} \in S^2$ we have
$$R(\boldsymbol{x}) \cdot R(\boldsymbol{e}_j) = \boldsymbol{x} \cdot \boldsymbol{e}_j = T(\boldsymbol{x}) \cdot T(\boldsymbol{e}_j) = T(\boldsymbol{x}) \cdot R(\boldsymbol{e}_j) \ .$$
Hence $R(\boldsymbol{x}) = T(\boldsymbol{x})$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Lemma 1.3 now shows that the isometries of S^2 are either the identity, rotations about axes through the origin, reflections in planes through the origin, or rotations about axes through the origin followed by reflection in the plane through the origin perpendicular to the axis (rotatory reflections).

2.3 Spherical Triangles

A *spherical triangle* Δ consists of three vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in S^2$ and three geodesics from $\boldsymbol{B}$ to $\boldsymbol{C}$, from $\boldsymbol{C}$ to $\boldsymbol{A}$, and from $\boldsymbol{A}$ to $\boldsymbol{B}$. The geodesics are the sides of Δ and their lengths will be denoted by a, b, and c respectively. If γ is a geodesics from $\boldsymbol{A}$ to $\boldsymbol{B}$, then γ is of the form
$$\gamma : [0, c] \to S^2 \ ; \quad t \mapsto (\cos t)\boldsymbol{A} + (\sin t)\boldsymbol{u}$$
where $\boldsymbol{u}$ is the unit tangent vector to the side $\boldsymbol{AB}$ of Δ at $\boldsymbol{A}$. Similarly, the unit tangent vector to the side $\boldsymbol{AC}$ at $\boldsymbol{A}$ is a vector $\boldsymbol{v}$. The angle between these unit vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is called the angle between the two sides $\boldsymbol{AB}$ and $\boldsymbol{AC}$ and is denoted by α. Similarly, we denote the angle between the two sides $\boldsymbol{BA}$ and $\boldsymbol{BC}$ at $\boldsymbol{B}$ is denoted by β and the angle between the two sides $\boldsymbol{CA}$ and $\boldsymbol{CB}$ at $\boldsymbol{C}$ is denoted by γ.
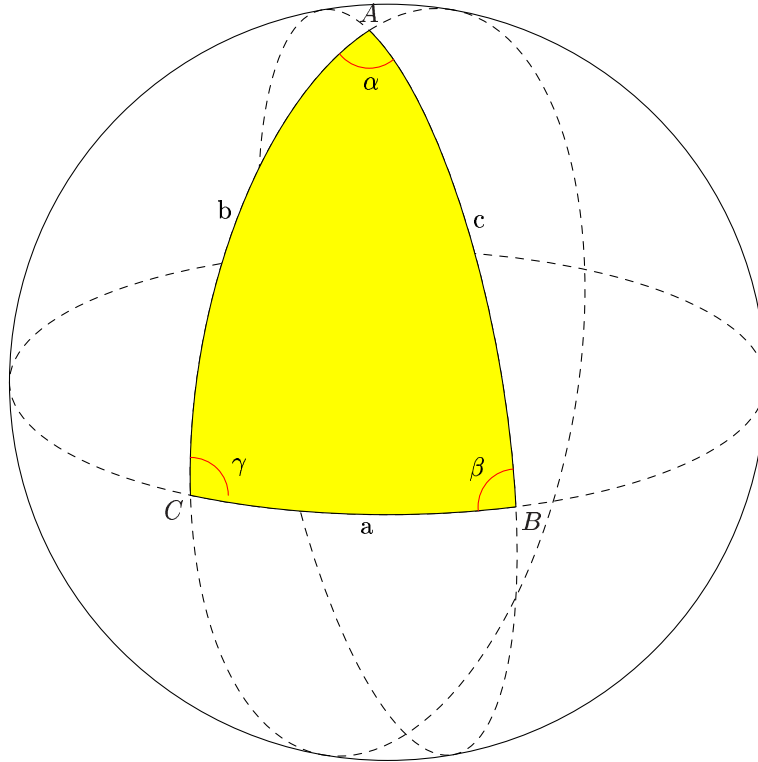
The sum of the angles of a spherical triangle is not π. Indeed the sum of the angles exceeds π by the area of the triangle.

Proposition 2.3 Gauss – Bonnet theorem for spherical triangles
For a triangle Δ on the unit sphere S^2 with area $\mathbb{A}(\Delta)$ we have

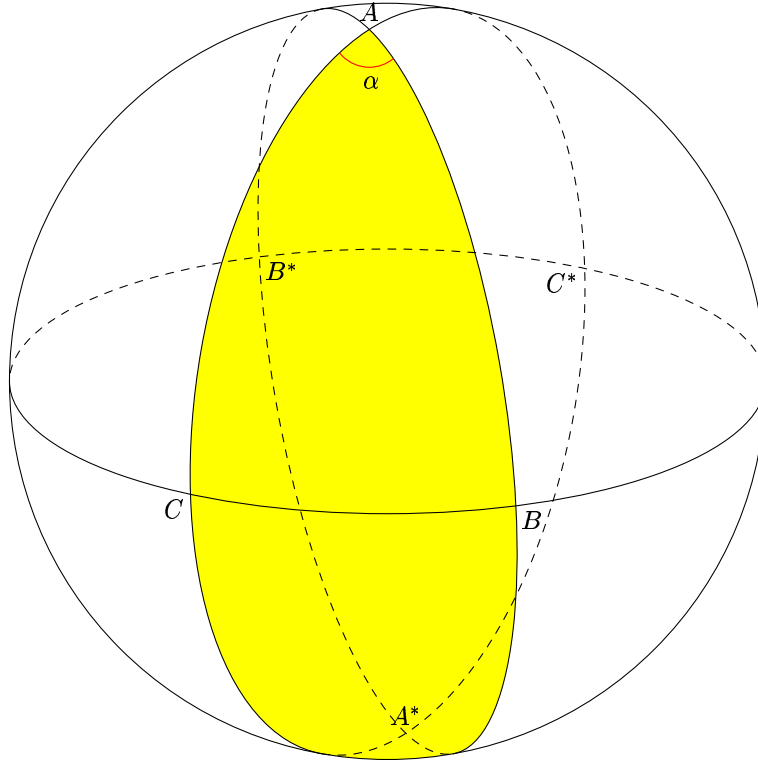$$\alpha + \beta + \gamma = \pi + \mathbb{A}(\Delta) \ .$$

Proof:

The three spherical lines that give the sides of Δ cross at $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and also at the antipodal points $\boldsymbol{A}^* = -\boldsymbol{A}, \boldsymbol{B}^* = -\boldsymbol{B}, \boldsymbol{C}^* = -\boldsymbol{C}$. These lines divide S^2 into 8 triangular regions. We will denote these by listing the vertices, so they are:

$$\Delta = \Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \ , \quad \Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C}) \ , \quad \Delta(\boldsymbol{A}, \boldsymbol{B}^*, \boldsymbol{C}) \ , \quad \Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}^*) \ ,$$
$$\Delta(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}) \ , \quad \Delta(\boldsymbol{A}, \boldsymbol{B}^*, \boldsymbol{C}^*) \ , \quad \Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C}^*) \ , \quad \Delta(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*) \ .$$

Since the mapping $J : \boldsymbol{x} \mapsto -\boldsymbol{x}$ is an isometry of S^2, we see that the area of the triangle $\Delta(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*)$ is equal to the area of $\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$. Similarly the areas of the other triangles are equal in pairs. Hence we see that

$$\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}^*, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}^*)) = \tfrac{1}{2}\mathbb{A}(S^2) = 2\pi \ . \quad (*)$$

The two triangles $\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ and $\Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C})$ together fill a sector of S^2 with angle α. So

$$\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C})) = \frac{\alpha}{2\pi}\mathbb{A}(S^2) = 2\alpha \ .$$

Similarly, we have

$$\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}^*, \boldsymbol{C})) = 2\beta$$
$$\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}^*)) = 2\gamma \ .$$

Adding these together gives

$$3\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + (\mathbb{A}(\Delta(\boldsymbol{A}^*, \boldsymbol{B}, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}^*, \boldsymbol{C})) + \mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}^*))) = 2(\alpha + \beta + \gamma) \ .$$

Now equation $(*)$ shows that this is $2\mathbb{A}(\Delta(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) + 2\pi$ so we have proved the proposition. □

Note that for triangles on a sphere of radius r the same argument shows that

$$\alpha + \beta + \gamma = \pi + \frac{\mathbb{A}(\Delta)}{r^2} \ .$$

We call $1/r^2$ the *curvature* of the sphere and denote it by K. So

$$\alpha + \beta + \gamma = \pi + K \, \mathbb{A}(\Delta) \ .$$

Suppose that we have a polygon P with N sides each of which is an arc of a spherical line. (We will only consider the case where N is at least 1 and the sides of the polygon do not cross one another, so P is simply connected.) If the internal angles of the polygon are $\theta_1, \theta_2, \ldots, \theta_N$, then we can divide it into $N - 2$ triangles and obtain

$$\theta_1 + \theta_2 + \ldots + \theta_N = (N - 2)\pi + \mathbb{A}(P) \ .$$

Now consider dividing the entire sphere into a finite number of polygonal faces by drawing arcs of spherical lines on the sphere. Let the number of polygonal faces be F, the number of arcs of spherical lines (edges) be E, and the number of vertices of the polygons V. The *Euler number* for this subdivision is $F - E + V$.

Proposition 2.4   Euler's formula for the sphere
Let the sphere be divided into F simply connected faces by drawing E arcs of spherical lines joining V vertices on S^2. Then

$$F - E + V = 2 \ .$$

Proof:
   Let the polygonal faces be $(P_j)_{j=1}^F$ and suppose that P_j has N_j edges. Then we have shown that the sum of the internal angles in P_j is $N_j\pi - 2\pi + \mathbb{A}(P_j)$. Summing over all of the faces we obtain:

$$\sum_{j=1}^{F}\sum \text{ angles in } P_j \ = \ \left(\sum_{j=1}^{F} N_j \pi\right) - \left(\sum_{j=1}^{F} 2\pi\right) + \left(\sum_{j=1}^{F} \mathbb{A}(P_j)\right) \ .$$

The angles at each vertex sum to 2π and each edge occurs in the boundary of two faces, so this gives

$$2\pi V = 2\pi E - 2\pi F + \mathbb{A}(S^2)$$

which simplifies to $F - E + V = 2$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

(It is essential that the polygonal faces are simply connected and indeed have at least one edge. Consider, for example, dividing the sphere into 3 faces by drawing edges approximating the tropics of Capricorn and Cancer. The number of edges and vertices are equal, so $F - E + V = 3$.)

The Cosine rule and Sine rule of Euclidean geometry have analogues for spherical triangles.

Proposition 2.5   Spherical Cosine Rule I
For a spherical triangle Δ

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha \ .$$

Proof:

We may apply an isometry to Δ to move $\boldsymbol{A}$ to the North pole $(0, 0, 1)$. Then a rotation about the z-axis will move the unit tangent vector $\boldsymbol{u}$ to $\boldsymbol{AB}$ at $\boldsymbol{A}$ to $(1, 0, 0)$. Then the point $\boldsymbol{B}$ is distance c from $\boldsymbol{A}$ in this direction, so $\boldsymbol{B} = (\sin c, 0, \cos c)$. The unit tangent vector $\boldsymbol{v}$ to $\boldsymbol{AC}$ at $\boldsymbol{A}$ must be at an angle α to $\boldsymbol{u}$, so it is $(\cos\alpha, \pm\sin\alpha, 0)$. By reflecting in the xz-plane, if necessary, we may ensure that $\boldsymbol{v} = (\cos\alpha, \sin\alpha, 0)$. The point $\boldsymbol{C}$ is at distance b from $\boldsymbol{A}$ in this direction, so $\boldsymbol{C} = (\cos b)\boldsymbol{A} + (\sin b)\boldsymbol{v} = (\sin b \cos\alpha, \sin b \sin\alpha, \cos b)$.

Now we can calculate the distance a from $\boldsymbol{B}$ to $\boldsymbol{C}$ directly.
$$\cos a = \boldsymbol{B} \cdot \boldsymbol{C} = (\sin c, 0, \cos c) \cdot (\sin b \cos\alpha, \sin b \sin\alpha, \cos b) = \sin b \sin c \cos\alpha + \cos b \cos c \ .$$

$\square$

Proposition 2.6   The Spherical Sine rule
For a spherical triangle Δ
$$\frac{\sin a}{\sin \alpha} = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma} \ .$$

Proof:

Let $\boldsymbol{C}^*$ be the reflection of $\boldsymbol{C}$ in the plane spanned by $\boldsymbol{A}$ and $\boldsymbol{B}$ (*i.e.* reflection in the spherical line through $\boldsymbol{A}$ and $\boldsymbol{B}$.) As in the proof of the cosine rule, we may assume that
$$\boldsymbol{A} = (0, 0, 1) \ , \quad \boldsymbol{B} = (\sin c, 0, \cos c) \ , \quad \boldsymbol{C} = (\sin b \cos\alpha, \sin b \sin\alpha, \cos b) \ .$$
Then $\boldsymbol{C}^* = (\sin b \cos\alpha, -\sin b \sin\alpha, \cos b)$.

The Euclidean distance $||\boldsymbol{C} - \boldsymbol{C}^*||$ between $\boldsymbol{C}$ and $\boldsymbol{C}^*$ is equal to $2\sin\frac{1}{2}d(\boldsymbol{C}, \boldsymbol{C}^*)$. So
$$2\sin\tfrac{1}{2}d(\boldsymbol{C}, \boldsymbol{C}^*) = 2\sin b \sin\alpha \ .$$

If we interchange $\boldsymbol{A}$ and $\boldsymbol{B}$ the point $\boldsymbol{C}$ is unchanged, so we must have
$$2\sin b \sin\alpha = 2\sin a \sin\beta \ .$$
Hence $\sin a / \sin\alpha = \sin b / \sin\beta$. $\square$

Recall that the *dual* of a vector space V over $\mathbb{R}$ is the vector space of all linear maps $\phi : V \to \mathbb{R}$. When V has an inner product we can identify a vector $\boldsymbol{v} \in V$ with the linear map $V \to \mathbb{R}; \ \boldsymbol{x} \mapsto \boldsymbol{v} \cdot \boldsymbol{x}$. This identifies the dual space V^* with V. This applies in particular to the space $\mathbb{R}^3$ with the usual inner product.

There is an interesting duality between the side lengths of a spherical triangle and the angles. Consider a spherical triangle Δ that does not have its three vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ linearly dependant. Then the vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ span a plane and there is a linear map $\mathbb{R}^3 \to \mathbb{R}$ in the dual space to $\mathbb{R}^3$ which has this plane as its kernel. Since we have an inner product on $\mathbb{R}^3$, we can find a unit vector A^* so that this linear map is $\boldsymbol{x} \mapsto \boldsymbol{A}^* \cdot \boldsymbol{x}$. Thus we have a unit vector $\boldsymbol{A}^*$ with
$$\boldsymbol{A}^* \cdot \boldsymbol{B} = \boldsymbol{A}^* \cdot \boldsymbol{C} = 0 \ .$$
There are just two choices for $\boldsymbol{A}^*$ and if we insist that $\boldsymbol{A}^*$ also satisfies $\boldsymbol{A}^* \cdot \boldsymbol{A} > 0$, then $\boldsymbol{A}^*$ is uniquely determined. We define the *dual triangle* Δ^* of Δ to be the spherical triangle with vertices $\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*$ satisfying:
$$\begin{array}{lll} \boldsymbol{A}^* \cdot \boldsymbol{A} > 0 \ ; & \boldsymbol{A}^* \cdot \boldsymbol{B} = 0 \ ; & \boldsymbol{A}^* \cdot \boldsymbol{C} = 0 \ ; \\ \boldsymbol{B}^* \cdot \boldsymbol{A} = 0 \ ; & \boldsymbol{B}^* \cdot \boldsymbol{B} > 0 \ ; & \boldsymbol{B}^* \cdot \boldsymbol{C} = 0 \ ; \\ \boldsymbol{C}^* \cdot \boldsymbol{A} = 0 \ ; & \boldsymbol{C}^* \cdot \boldsymbol{B} = 0 \ ; & \boldsymbol{C}^* \cdot \boldsymbol{C} > 0 \ . \end{array}$$

Note that the unit vector $\boldsymbol{A}$ satisfies $\boldsymbol{A} \cdot \boldsymbol{B}^* = \boldsymbol{A} \cdot \boldsymbol{C}^* = 0$ and $\boldsymbol{A} \cdot \boldsymbol{A}^* > 0$, so the original triangle Δ is the dual triangle of Δ^*. The vertices of Δ^* represent the sides of Δ and the sides of Δ^* represent the vertices of Δ.

We now wish to find how the angles and side lengths of Δ^* are related to those of Δ.

Proposition 2.7   Dual spherical triangles

Let Δ be a spherical triangle with angles α, β, γ and side lengths a, b, c. Then the dual triangle Δ^* has sides of length $a^* = \pi - \alpha, b^* = \pi - \beta, c^* = \pi - \gamma$ and angles $\alpha^* = \pi - a, \beta^* = \pi - b, \gamma^* = \pi - c$.

Proof:

We may apply an isometry to Δ so that $\boldsymbol{A} = (1, 0, 0)$ and the unit tangent vectors to $\boldsymbol{AB}$ and $\boldsymbol{AC}$ at $\boldsymbol{A}$ are $\boldsymbol{u} = (1, 0, 0)$ and $\boldsymbol{v} = (\cos \alpha, \sin \alpha, 0)$ respectively. Then $\boldsymbol{C}^*$ is perpendicular to the plane through $\boldsymbol{A}$ and $\boldsymbol{B}$, so it is perpendicular to $\boldsymbol{A} = (0, 0, 1)$ and $\boldsymbol{u} = (1, 0, 0)$. Hence, $\boldsymbol{C}^* = \pm(0, 1, 0)$. Since we also want $\boldsymbol{C}^* \cdot \boldsymbol{B} > 0$ we must have $\boldsymbol{C}^* = (0, 1, 0)$. Similarly, $\boldsymbol{B}^* = (\sin \alpha, -\cos \alpha, 0)$. Consequently, $\boldsymbol{B}^* \cdot \boldsymbol{C}^* = -\cos \alpha$ and so the distance from $\boldsymbol{B}^*$ to $\boldsymbol{C}^*$ is $\pi - \alpha$.
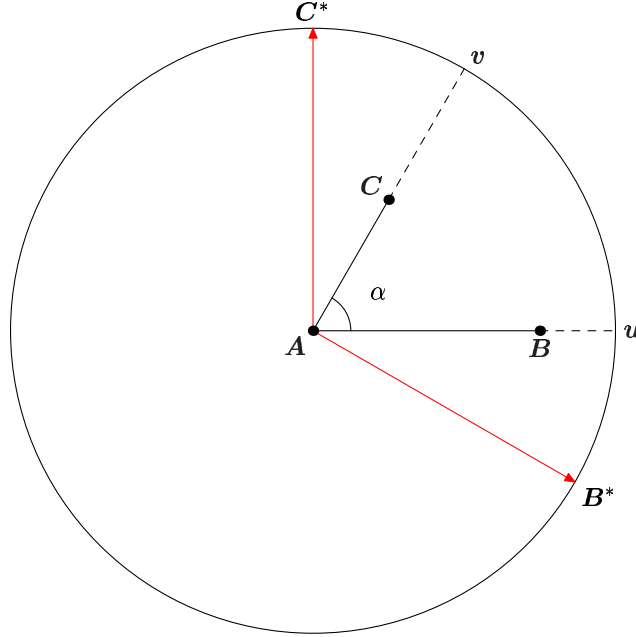
By interchanging the rôles of $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ in the above argument we see that the side lengths a^*, b^*, c^* in Δ^* are

$$a^* = \pi - \alpha \;, \qquad b^* = \pi - \beta \;, \qquad c^* = \pi - \gamma \;.$$

The triangle Δ is also the dual triangle to Δ^*, so the angles $\alpha^*, \beta^*, \gamma^*$ in Δ^* must satisfy

$$a = \pi - \alpha^* \;, \qquad b = \pi - \beta^* \;, \qquad c = \pi - \gamma^* \;.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

By applying the Cosine rule to the dual triangle we obtain a second form of the Cosine rule. This form enables us to calculate the side lengths of a spherical triangle when we know the angles. Unlike the Euclidean situation, the angles of a spherical triangle determine the triangle up to isometry.

Corollary 2.8   Spherical Cosine Rule II

For a spherical triangle Δ

$$\cos \alpha = -\cos \beta \cos \gamma + \sin \beta \sin \gamma \cos a \;.$$

16

Proof:

When we apply the Cosine rule of Proposition 2.5 to the dual triangle Δ^* we obtain

$$\cos a^* = \cos b^* \cos c^* + \sin b^* \sin c^* \cos \alpha^* \ .$$

The previous proposition shows that this is equivalent to

$$-\cos \alpha = (-\cos \beta)(-\cos \gamma) + \sin \beta \sin \gamma (-\cos a) \ .$$

$\square$

2.4 *The projective plane*

In the Renaissance artists thought about perspective drawings. An object was projected onto a plane canvas by constructing a line from each point of the object to the eye and marking where it crossed the plane. Of course, moving the plane of the canvas changed the drawing. Instead of thinking about the points on the plane, we could think about the straight lines to the eye. Most of these lines will cross a fixed plane at just one point but, if the line happens to be parallel to the plane, the line does not meet the plane at all. Mathematicians abstracted this idea to define the projective plane.

The (real) projective plane $\mathbb{RP}^2$ is the set of all 1-dimensional vector subspaces of $\mathbb{R}^3$. For any point $(x, y, z) \in \mathbb{R}^3$ except $(0, 0, 0)$, there is an unique 1-dimensional vector subspace containing it. We write this subspace as

$$[x : y : z] = \{\lambda(x, y, z) : \lambda \in \mathbb{R}\} \ .$$

So $[x : y : z]$ is a point in the projective plane, provided that x, y and z are not all 0. Note that $[x : y : z] = [\mu x : \mu y : \mu z]$ for any non-zero scalar $\mu \in \mathbb{R}$. We call the numbers x, y, z the *homogeneous co-ordinates* of $[x : y : z]$.

We can think of the projective plane as being an Euclidean plane together with a circle of extra points adjoined "at infinity". For let Π be the plane $\{(x, y, z) : z = 1\}$ in $\mathbb{R}^3$. This is an Euclidean plane lying inside $\mathbb{R}^3$ but not containing the origin. A point $[x : y : z]$ in the projective plane is a 1-dimensional vector subspace of $\mathbb{R}^3$. This subspace crosses Π at an unique point $(x/z, y/z, 1)$ provided that $z \neq 0$. When $z = 0$ the 1-dimensional subspace is parallel to Π. We can therefore identify each point $(x, y, 1) \in \Pi$ with the point $[x : y : 1] \in \mathbb{RP}^2$. The remaining points of $\mathbb{RP}^2$ are

$$\{[x : y : 0] : x, y \text{ not both } 0\} \ .$$

There is one point for each 1-dimensional vector subspace of $\mathbb{R}^3$ parallel to Π. Changing the plane Π changes which points of $\mathbb{RP}^2$ are in the plane and which points are the extra ones.

Let ℓ be a 2-dimensional vector subspace of $\mathbb{R}^3$. Then the set of all the 1-dimensional vector subspaces of ℓ is a subset $\{\boldsymbol{P} \in \mathbb{RP}^2 : \boldsymbol{P} \subset \ell\}$ of the projective plane. We will call it a *line* in $\mathbb{RP}^2$ and also denote it by ℓ. Each 2-dimensional vector subspace of $\mathbb{R}^3$ is of the form

$$\ell = \{(x, y, z) \in \mathbb{R}^3 : ax + by + cz = 0\}$$

for some numbers a, b, c not all of which are 0. This subspace ℓ intersects Π in the Euclidean straight line $\{(x, y, 1) : ax + by + c = 0\}$ provided that a and b are not both 0. In addition, there is the extra point $[b : -a : 0] \in \mathbb{RP}^2$. When a and b are both 0, we have $\ell = \{(x, y, z) : z = 0\}$ and this gives all the extra points $[x : y : 0]$ in the projective plane.

Suppose that ℓ_1, ℓ_2 are two different lines in $\mathbb{RP}^2$, so ℓ_1, ℓ_2 are 2-dimensional vector subspaces of $\mathbb{R}^3$. Then the intersection $\ell_1 \cap \ell_2$ is a 1-dimensional vector subspace of $\mathbb{R}^3$. Therefore the two lines ℓ_1, ℓ_2 always meet at a single point of $\mathbb{RP}^2$.

Two different Euclidean straight lines m_1, m_2 correspond to two projective lines. If m_1 and m_2 meet at a point in Π, then the projective lines will meet at the corresponding point in $\mathbb{RP}^2$. However, if m_1 and m_2 are parallel, then the projective lines will meet at one of the extra points $[x : y : 0] \in \mathbb{RP}^2$. Hence, we may think of the projective plane as a copy of the Euclidean plane Π together with a set of additional points $[x : y : 0]$, one added for each set of parallel lines in Π. These points are added "at infinity" to provide points where parallel lines in Π meet.

If we wish to put a metric on the projective plane, it is better to use a different representation. Each 1-dimensional vector subspace of $\mathbb{R}^3$ cuts the unit sphere S^2 at two antipodal points. So the projective plane $\mathbb{RP}^2$ as the set of pairs of antipodal points in S^2. We can therefore think of the projective plane as being the space we obtain by taking the upper hemisphere $\{(x, y, z) \in S^2 : z \geqslant 0\}$ of S^2 and identifying the boundary points on the equator by identifying $(x, y, 0)$ with the antipodal point $(-x, -y, 0)$.

More formally, we define a quotient map

$$q : S^2 \to \mathbb{RP}^2 ; \quad (x, y, z) \mapsto [x : y : z] .$$

This map is 2-to-1. The spherical lines in S^2 are mapped by q onto projective lines in $\mathbb{RP}^2$. We can use the metric on the sphere to give a metric on the projective plane. Define the distance between two points $\boldsymbol{A}$ to $\boldsymbol{B}$ of $\mathbb{RP}^2$ to be

$$d(\boldsymbol{A}, \boldsymbol{B}) = \inf\{d(\boldsymbol{x}, \boldsymbol{y}) : q(\boldsymbol{x}) = \boldsymbol{A} \ \text{ and } \ q(\boldsymbol{y}) = \boldsymbol{B} \} .$$

It is easy to see that this defines a metric. Moreover, if $\boldsymbol{A}, \boldsymbol{B}$ are distinct points, there is an unique projective line through them. One arc of this line gives the shortest path from $\boldsymbol{A}$ to $\boldsymbol{B}$ in $\mathbb{RP}^2$, which is of length $d(\boldsymbol{A}, \boldsymbol{B})$. This is the unique geodesic from $\boldsymbol{A}$ to $\boldsymbol{B}$. As in the sphere, we define the angle between two such geodesics from $\boldsymbol{A}$ to points $\boldsymbol{B}$ and $\boldsymbol{C}$ to be the angle between the corresponding unit tangent vectors at $\boldsymbol{A}$.

We can now consider triangles Δ in the projective plane. These have three distinct vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Between each pair of vertices, there is an unique geodesic forming a side of Δ. The Cosine rule and Sine rule for spherical triangles readily give the corresponding rules for triangles in the projective plane.

The geometry of the projective plane $\mathbb{RP}^2$ is very similar to that of the sphere S^2. Indeed, the geometry of the projective plane is simpler because we do not need to worry about the exceptional cases where two points of the sphere are antipodal.

3. STEREOGRAPHIC PROJECTION

3.1 Definition

In the Algebra and Geometry course you met stereographic projection. This is a map $\pi : S^2 \to \mathbb{C}_\infty$ that we use to identify the sphere with the extended complex plane $\mathbb{C}_\infty$. We think of the complex plane as the plane $\{x_3 = 0\}$ in $\mathbb{R}^3$ and project a point $\boldsymbol{a} = (a_1, a_2, a_3) \in S^2$ from the North pole $(0, 0, 1)$ to the point $(u, v, 0)$ that corresponds to $w = u + iv \in \mathbb{C}_\infty$. So $(0, 0, 1)$, $\boldsymbol{a}$ and $(u, v, 0)$ all lie on a straight line. Thus there is a $t \in \mathbb{R}$ with

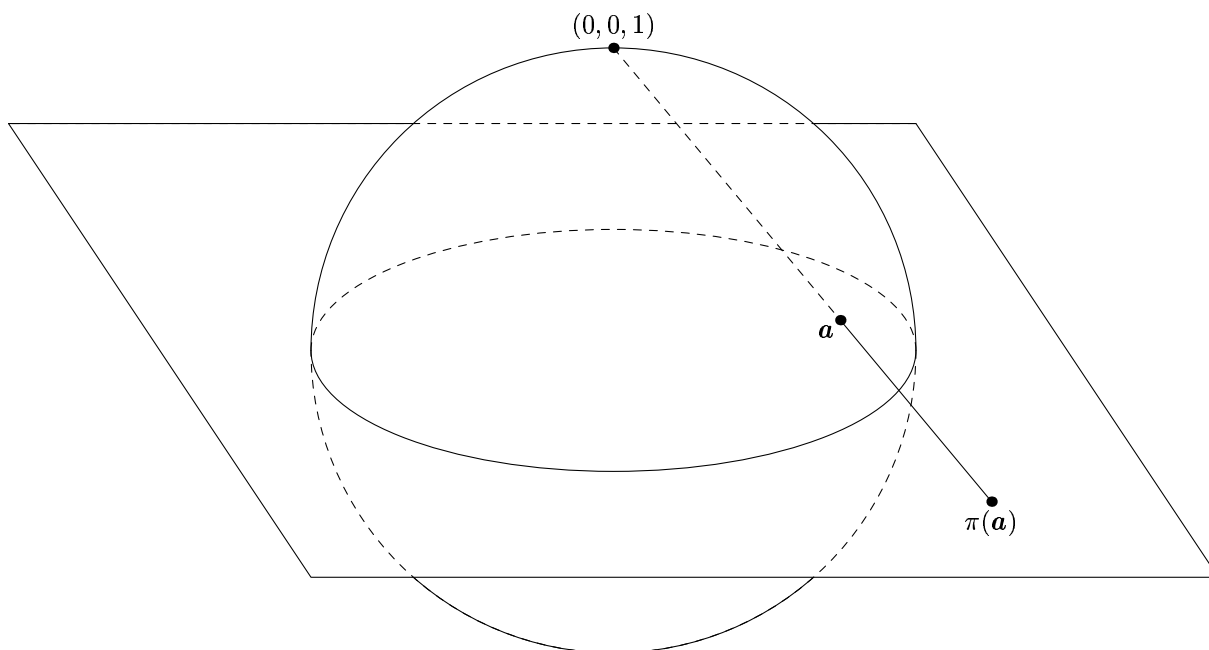$$(u, v, 0) = t(a_1, a_2, a_3) + (1 - t)(0, 0, 1) \ .$$

Solving this we get

$$(u, v) = \left(\frac{a_1}{1 - a_3}, \frac{a_2}{1 - a_3} \right) \ .$$

In a similar way we can find the inverse map $\pi^{-1} : \mathbb{C}_\infty \to S^2$:

$$\boldsymbol{a} = \pi^{-1}(w) = \left(\frac{2u}{1 + |w|^2}, \frac{2v}{1 + |w|^2}, \frac{|w|^2 - 1}{1 + |w|^2} \right) \ .$$

We should think of the complex numbers as giving co-ordinates on the sphere. Compare this with the projections used to produce flat maps in atlases.

The metric on the sphere gives a metric on the extended complex plane. We will always use this metric on the extended complex plane.

Proposition 3.1 The spherical distance on $\mathbb{C}_\infty$.
If $\boldsymbol{a}, \boldsymbol{b} \in S^2$ project to $w, z \in \mathbb{C}_\infty$, we define $d(w, z)$ to be the spherical distance $d(\boldsymbol{a}, \boldsymbol{b})$. Then

$$\sin \tfrac{1}{2} d(w, z) = \frac{|w - z|}{\sqrt{1 + |w|^2}\sqrt{1 + |z|^2}} \ .$$

(Note that when, say, $z = \infty$ we should interpret this formula as

$$\sin \tfrac{1}{2} d(w, \infty) = \frac{1}{\sqrt{1 + |w|^2}} \ .)$$

Proof:

Write $w = u + iv, z = x + iy$ and apply the formulae above for stereographic projection:

$$\cos d(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} \cdot \boldsymbol{b} = \left(\frac{2u}{1 + |w|^2}, \frac{2v}{1 + |w|^2}, \frac{|w|^2 - 1}{1 + |w|^2} \right) \cdot \left(\frac{2x}{1 + |z|^2}, \frac{2y}{1 + |z|^2}, \frac{|z|^2 - 1}{1 + |z|^2} \right)$$

$$= \frac{4ux + 4vy + (|w|^2 - 1)(|z|^2 - 1)}{(1 + |w|^2)(1 + |z|^2)}$$

$$= 1 - \frac{2|w - z|^2}{(1 + |w|^2)(1 + |z|^2)}$$

and this gives the desired result. $\qquad \square$

Note that the maximum distance between two points $w, z \in \mathbb{C}_\infty$ is π. This occurs when the corresponding points of S^2 are antipodal, that is when $z = -1/\overline{w}$.

3.2 Möbius transformations

We can use the last Proposition to identify the isometries of S^2 with maps from $\mathbb{C}_\infty$ to itself.

Proposition 3.2 Möbius transformations as isometries of $\mathbb{C}_\infty$
The isometries of $\mathbb{C}_\infty$, with the spherical metric, are the Möbius transformations

$$T : z \mapsto \frac{az + b}{-\overline{b} z + \overline{a}} \qquad \text{where} \quad |a|^2 + |b|^2 = 1$$

and their conjugates

$$\overline{T} : z \mapsto \overline{\left(\frac{az + b}{-\overline{b} z + \overline{a}} \right)} \qquad \text{where} \quad |a|^2 + |b|^2 = 1 \ .$$

Proof:

The map T satisfies

$$\sin \tfrac{1}{2} d(T(w), T(z)) = \frac{\left| \left(\frac{aw + b}{-\overline{b}w + \overline{a}} \right) - \left(\frac{az + b}{-\overline{b}z + \overline{a}} \right) \right|}{\sqrt{1 + \left| \frac{aw + b}{-\overline{b}w + \overline{a}} \right|^2} \sqrt{1 + \left| \frac{az + b}{-\overline{b}z + \overline{a}} \right|^2}}$$

$$= \frac{|(aw + b)(-\overline{b}z + \overline{a}) - (az + b)(-\overline{b}w + \overline{a})|}{\sqrt{|-\overline{b}w + \overline{a}|^2 + |aw + b|^2} \sqrt{|-\overline{b}z + \overline{a}|^2 + |az + b|^2}}$$

$$= \frac{|w - z|}{\sqrt{1 + |w|^2} \sqrt{1 + |z|^2}} = \sin \tfrac{1}{2} d(w, z) \ .$$

So it is an isometry. Since conjugation is clearly an isometry, it follows that $\overline{T}$ is also an isometry. It remains to show that these are the only isometries of $\mathbb{C}_\infty$.

Suppose that $R : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is an isometry and set $z_o = R(0)$. Then we can choose $a, b \in \mathbb{C}$ with $z_o = a/\overline{b}$ and $|a|^2 + |b|^2 = 1$. This gives an isometry $T : z \mapsto (az + b)/(-\overline{b}z + \overline{a})$ with $T(0) = z_o$. Hence $U = T^{-1} \circ R$ is an isometry that fixes 0. Now, under stereographic projection, $0 \in \mathbb{C}_\infty$ corresponds to the South pole $(0, 0, -1) \in S^2$. So U must correspond to either a rotation of S^2 about the x_3-axis or to a reflection in a plane through this axis. It is easy to see that these isometries are given by either

$$U : z \mapsto e^{2i\theta} z \qquad \text{or} \qquad U : z \mapsto e^{2i\theta} \overline{z} \ .$$

Therefore, $R = T \circ U$ is either a Möbius transformation of the required form (if U is a rotation) or the conjugate of a Möbius transformation (if U is a reflection). $\qquad \square$

The transformations T in this Proposition are orientation preserving and the transformations $\overline{T}$ are orientation reversing.

Note that the orientation preserving isometries of $\mathbb{C}_\infty$ are the Möbius transformations

$$z \mapsto \frac{az+b}{cz+d} \quad \text{for which the matrix} \quad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is unitary with determinant 1, that is $M \in \mathrm{SU}(2)$. Two such matrices M, M' define the same Möbius transformation only when $M' = \pm M$. So we see that the group of orientation preserving isometries of $\mathbb{C}_\infty$ is $\mathrm{SU}(2)/\{\pm I\}$. This is called the *projective special unitary group* $\mathrm{PSU}(2)$. We know that the orientation preserving isometries of S^2 are $\mathrm{SO}(3)$. So stereographic projection gives an isomorphism between the groups

$$\mathrm{SO}(3) \quad \text{and} \quad \mathrm{PSU}(2) = \frac{\mathrm{SU}(2)}{\{\pm I\}} \ .$$

Finally, the matrices $M \in \mathrm{SU}(2)$ are those of the form

$$\begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \quad \text{with} \quad |a|^2 + |b|^2 = 1 \ .$$

These correspond to the pair $(a, b) \in \mathbb{C}^2$ that lie in the unit sphere S^3. So $\mathrm{SU}(2)$ can be identified with the 3-dimensional sphere S^3. The quotient $\mathrm{PSU}(2) = \mathrm{SU}(2)/\{\pm I\}$ is then identified with the set of pairs of antipodal points in S^3, that is the projective 3-space $\mathbb{RP}^3$ consisting of all 1-dimensional vector subspaces of $\mathbb{R}^4$.

3.3 Riemannian metrics

Now consider the lengths of curves. For a curve $\gamma : [0, 1] \to S^2$ we can write

$$\gamma(t) = (x_1(t), x_2(t), x_3(t)) \ .$$

Then its length is

$$L(\gamma) = \int_0^1 \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2 + \left(\frac{dx_3}{dt}\right)^2} \, dt \ .$$

It is convenient to abbreviate this expression by writing ds for the length of an infinitesimal displacement (dx_1, dx_2, dx_3). Then

$$L(\gamma) = \int_\gamma ds \quad \text{where} \quad ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \ .$$

We say that the *Riemannian metric* on S^2 is given by

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \ .$$

We want a similar formula for lengths of curves $\beta : [0, 1] \to \mathbb{C}_\infty \ ; \quad t \mapsto w = u + iv$ on $\mathbb{C}_\infty$. Then $\gamma = \pi^{-1} \circ \beta$ is the corresponding curve on S^2 so the length of β is

$$L(\beta) = L(\gamma) = \int_0^1 ds \ .$$

Now, for infinitesimal displacements, the formula in Proposition 3.1 becomes

$$\tfrac{1}{2} ds = \frac{|dw|}{1 + |w|^2} \ .$$

21

So we have

$$L(\beta) = \int_\beta \frac{2}{1 + |w|^2} \, |dw| = \int_0^1 \left(\frac{2}{1 + |w|^2} \right) \left| \frac{dw}{dt} \right| \, dt \; .$$

So the Riemannian metric on $\mathbb{C}_\infty$ is

$$ds^2 = \frac{2}{1 + |w|^2} \, |dw|^2 = \left(\frac{2}{1 + u^2 + v^2} \right) (du^2 + dv^2) \; .$$

We could have used this Riemannian metric on $\mathbb{C}_\infty$ to define the lengths of curves and hence done all our spherical geometry on $\mathbb{C}_\infty$.

Note that the Riemannian metric ds^2 on $\mathbb{C}_\infty$ is just a scaling of the Euclidean metric $du^2 + dv^2$ by $(2/(1 + |w|^2))^2$. This means that lengths are changed from the Euclidean metric but angles are not. Thus stereographic projection is conformal.

4. THE HYPERBOLIC PLANE

In the last section we considered a Riemannian metric $ds = \dfrac{2|dz|}{1 + |z|^2}$ on $\mathbb{C}_\infty$ that made it isometric to the unit sphere. In this section we want to study an even more important Riemannian metric called the hyperbolic metric. We will begin by defining the Riemannian metric directly and studying its properties.

4.1 Poincaré's disc model

Our first model for the hyperbolic plane will be the unit disc $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. A *hyperbolic line* for this model is the intersection with $\mathbb{D}$ of a circle, or a straight line, that crosses the boundary of $\mathbb{D}$ orthogonally. There is a large group of symmetries of the hyperbolic plane, as the next proposition shows.

Proposition 4.1 Möbius transformations of $\mathbb{D}$
A Möbius transformation T maps $\mathbb{D}$ onto itself if and only if T is of the form

$$T : w \mapsto \frac{aw + b}{\overline{b}w + \overline{a}}$$

for some $a, b \in \mathbb{C}$ with $|a|^2 - |b|^2 = 1$. In this case T maps each hyperbolic line in $\mathbb{D}$ onto a hyperbolic line.

Proof:

Suppose that T is of the form given above. If $|w| = 1$ then

$$|T(w)| = \left| \frac{aw + b}{\overline{b}w + \overline{a}} \right| = \frac{|aw + b|}{|w||\overline{aw + b}|} = 1 \ .$$

Now a Möbius transformation maps a circle to a circle or a straight line, so T must map the unit circle onto itself. Hence T maps $\mathbb{D}$ either onto $\mathbb{D}$ or onto $\{z \in \mathbb{C}_\infty : |z| > 1\}$. Since $|T(0)| = |b/\overline{a}| < 1$, it must map the $\mathbb{D}$ onto itself. Furthermore T is conformal so it preserves angles. Hence T maps a circle (or straight line) orthogonal to $\partial \mathbb{D}$ to another such circle or straight line. This means that T maps a hyperbolic line in $\mathbb{D}$ onto another hyperbolic line.

Conversely, suppose that T is any Möbius transformation that maps $\mathbb{D}$ onto itself. Then T must map the boundary $\partial \mathbb{D}$ of $\mathbb{D}$ onto itself. Let $z^* = 1/\overline{z}$ be the point inverse to z in the unit circle. Then $T(z^*)$ is the point inverse to $T(z)$ in $T(\partial \mathbb{D}) = \partial \mathbb{D}$. Hence, $T(z^*)^* = T(z)$. We may assume that T is of the form

$$T : z \mapsto \frac{az + b}{cz + d} \qquad \text{where} \quad ad - bc = 1 \ .$$

Then

$$T(z^*)^* = \frac{\overline{c} + \overline{d}z}{\overline{a} + \overline{b}z}$$

for each $z \in \mathbb{C}_\infty$. This means that the two matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} \overline{d} & \overline{c} \\ \overline{b} & \overline{a} \end{pmatrix}$ give the same Möbius transformation, so they must be equal up to a constant multiple. Both matrices have determinant 1, so the multiple must be ± 1. Thus we have $c = \pm \overline{b}$ and $d = \pm \overline{a}$.

Since $ad - bc = 1$ we must have $\pm(|a|^2 - |b|^2) = 1$. Also $T(0) = \pm b/\overline{a}$ lies in $\mathbb{D}$, so $|b| < |a|$. Hence we must have the $+$ sign in the formulae. Therefore,

$$T(z) = \frac{az + b}{\overline{b}z + \overline{a}} \qquad \text{with} \quad |a|^2 - |b|^2 = 1 \ .$$

$\square$

We will denote the set of all the Möbius transformations that map $\mathbb{D}$ onto $\mathbb{D}$ by $\text{Aut}^+(\mathbb{D})$. The $+$ sign is to remind us that each of these transformations is orientation preserving. Reflection in the x-axis $R : z \mapsto \overline{z}$ is an orientation reversing map from $\mathbb{D}$ onto $\mathbb{D}$ that maps hyperbolic lines to hyperbolic lines. More generally, for each $T \in \text{Aut}^+(\mathbb{D})$, the map

$$\overline{T} : z \mapsto \overline{T(z)} = \frac{\overline{a}z + \overline{b}}{\overline{b}\overline{z} + a}$$

is an orientation reversing map from $\mathbb{D}$ onto $\mathbb{D}$ that maps hyperbolic lines to hyperbolic lines. These form a set $\text{Aut}^-(\mathbb{D})$ and we write $\text{Aut}(\mathbb{D})$ for the union $\text{Aut}^+(\mathbb{D}) \cup \text{Aut}^-(\mathbb{D})$. It is clear that $\text{Aut}(\mathbb{D})$ is a group under composition and that the orientation preserving maps $\text{Aut}^+(\mathbb{D})$ form a normal subgroup of index 2. This group is in fact the group of all isometries of the hyperbolic plane.

Let us consider some examples of symmetries of $\mathbb{D}$. First consider those symmetries that fix 0. Each rotation $z \mapsto e^{i\theta}z$ about 0 is in $\text{Aut}^+(\mathbb{D})$ and each reflection in a line through the origin $z \mapsto e^{i\theta}\overline{z}$ is in $\text{Aut}^-(\mathbb{D})$. There are many other symmetries that move 0 to another point of $\mathbb{D}$. Indeed, for each point $z_o \in \mathbb{D}$ we can find a, b with $z_o = b/\overline{a}$ and $|a|^2 - |b|^2 = 1$. Then the Möbius transformation T maps 0 to z_o. We can take $a = 1/\sqrt{1 - |z_o|^2}$ and $b = z_o/\sqrt{1 - |z_o|^2}$, and so get

$$T : w \mapsto \frac{w + z_o}{1 + \overline{z_o}w} \ .$$

Thus we can always find a symmetry $T \in \text{Aut}^+(\mathbb{D})$ that sends 0 to any given point in $\mathbb{D}$. This shows that the geometry of the hyperbolic plane is the same at each point. The origin $0 \in \mathbb{D}$ looks special for the Euclidean geometry of the disc but it is not special for hyperbolic geometry.

Let ℓ be a hyperbolic line in $\mathbb{D}$. Then ℓ is an arc of a circle orthogonal to $\partial\mathbb{D}$. Choose a point z_o on ℓ and let T be the symmetry

$$T : w \mapsto \frac{w + z_o}{1 + \overline{z_o}w} \ .$$

Then $T^{-1}(\ell)$ is a hyperbolic line through $T^{-1}(z_o) = 0$ and is therefore a diameter δ through 0. Reflection in δ is an orientation reversing symmetry R of $\mathbb{D}$. Hence the conjugate $S = T \circ R \circ T^{-1}$ is an orientation reversing symmetry of $\mathbb{D}$. It is clear that S fixes each point of ℓ. We call it the *reflection in the hyperbolic line* ℓ. We can also think of it as inversion in the circle containing ℓ. For the points z and $R(z)$ are inverse in the straight line containing δ, so $w = T(z)$ and $S(w) = T(R(z))$ are inverse in the circle containing ℓ. Thus reflection in ℓ is the same as inversion in the circle containing ℓ.

We can use the group of symmetries of the hyperbolic plane to simplify arguments. We will do this throughout the remainder of this section. As a first example we have:

Proposition 4.2   Hyperbolic lines
For any two distinct points z_0, z_1 in the disc $\mathbb{D}$, there is an unique hyperbolic line passing through them.

Proof:
      Consider first the case when $z_0 = 0$. Then there is certainly a diameter of the disc that passes through z_1 and this is one hyperbolic line through z_0 and z_1. Suppose that C is a circle (or straight line) that crosses the boundary of $\mathbb{D}$ orthogonally and passes through $0 = z_0$ and z_1. Then inversion in the boundary of $\mathbb{D}$ must map C onto itself. Hence C must pass through the point ∞ that is inverse to $0 = z_0$. This means that C must be a straight line through $0 = z_0$ and z_1. There is only one such line. Thus we have proven the result when $z_0 = 0$.

Now consider the general case. Let T be the Möbius transformation

$$T : z \mapsto \frac{z + z_0}{1 + \overline{z_0}z}$$

in $\text{Aut}^+(\mathbb{D})$. Then T sends 0 to z_0 and the point $w = (z_1 - z_0)/(1 - \overline{z_0}z_1)$ to z_1. Thus T maps the hyperbolic line through 0 and w to a hyperbolic line through z_0 and z_1. Since we have already established the result for the points 0 and w, the proof is complete. $\square$

We often abbreviate the latter part of this argument to: *"By applying a symmetry of $\mathbb{D}$ we may assume that the point z_0 is 0"*.

Let ℓ be a hyperbolic line in $\mathbb{D}$ and $\boldsymbol{P}$ a point of $\mathbb{D}$ that does not lie on ℓ. By applying a symmetry we may assume that $\boldsymbol{P} = 0$ and then the hyperbolic lines through $\boldsymbol{P}$ are the diameters of $\mathbb{D}$. It is now clear that there are an infinite number of hyperbolic lines m that pass through $\boldsymbol{P}$ but do not meet ℓ. The extreme case is when ℓ and m have a boundary point in common. In this case we say that ℓ and m are *parallel*. Otherwise ℓ and m do not meet either in $\mathbb{D}$ or on the boundary and we say that they are *ultraparallel*. The three geometries that we have studied differ markedly in this respect. In the Euclidean plane, there is an unique line through a point $\boldsymbol{P}$ not on a line ℓ that is disjoint from ℓ. This is the line through $\boldsymbol{P}$ parallel to ℓ. On the sphere, there are no such lines since every pair of spherical lines intersect. In the hyperbolic plane, there are several such lines.
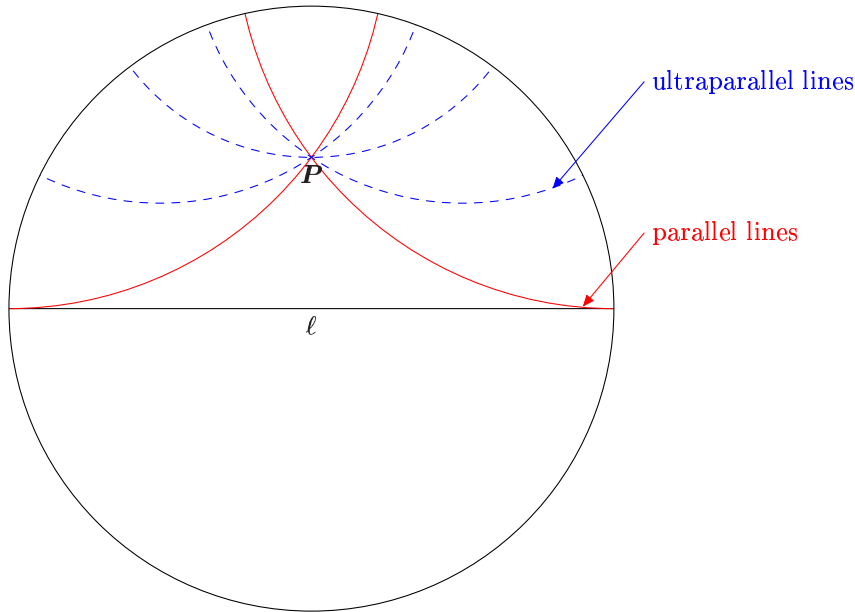
We now want to define a Riemannian metric on $\mathbb{D}$ so that we can define hyperbolic lengths and angles. We want a metric that is invariant under each of the symmetries in $\mathrm{Aut}(\mathbb{D})$. The *hyperbolic Riemannian metric on $\mathbb{D}$* is given by
$$ds = \frac{2|dz|}{1 - |z|^2} \ .$$
This means that a smooth curve $\gamma : [a, b] \to \mathbb{D}$ has hyperbolic length
$$L(\gamma) = \int_a^b ds = \int_a^b \frac{2}{1 - |\gamma(t)|^2} \, |\gamma'(t)| \, dt \ .$$

Let δ be an infinitesimally small number. Then the line segment from z to $z + \frac{1}{2}(1 - |z|^2)\delta$ has hyperbolic length $|\delta|$. In other words a measuring rod of constant hyperbolic length appears to our Euclidean eyes to shrink as it approaches the boundary of $\mathbb{D}$. Angles in the hyperbolic geometry of $\mathbb{D}$ are the same as Euclidean angles. For suppose that we wish to measure the angle between two curves that meet at a point $z_o \in \mathbb{D}$. To do this we only need to look close to z_o and here the metric $ds = 2|dz|/(1 - |z|^2)$ is virtually constant at $2/(1 - |z_o|^2)$ times the Euclidean metric. Since rescaling lengths does not change angles, this means that the Euclidean angle between the curves at z_o and the hyperbolic angle are the same.

We can define the *hyperbolic distance $\rho(z_0, z_1)$ between two points $z_0, z_1 \in \mathbb{D}$* to be the infimum of the hyperbolic length of all smooth curves from z_0 to z_1 in $\mathbb{D}$.

Proposition 4.3 The hyperbolic metric on the disc
The hyperbolic metric $\rho(\cdot, \cdot)$ is a metric on $\mathbb{D}$.

Proof:

First observe that the hyperbolic metric has

$$ds = \frac{2|dz|}{1 - |z|^2} \geqslant 2|dz| \ .$$

Hence, the hyperbolic length of a curve $\gamma : [a, b] \to \mathbb{D}$ from z_0 to z_1 satisfies

$$L(\gamma) = \int_a^b ds \geqslant \int_a^b 2 \, |dz| = 2 \int_a^b |\gamma'(t)| \, dt \geqslant 2|\gamma(b) - \gamma(a)| = 2|z_1 - z_0| \ .$$

This certainly shows that $\rho(z_0, z_1) \geqslant 2|z_1 - z_0| \geqslant 0$ with equality only when $z_1 = z_0$.

If γ is a curve from z_0 to z_1, then its reverse is a curve from z_1 to z_0 with the same length. Hence $\rho(z_0, z_1) = \rho(z_1, z_0)$.

Finally, if β is a curve from z_0 to z_1 and γ is a curve from z_1 to z_2, then β followed by γ is a curve δ from z_0 to z_1. Since $L(\delta) = L(\beta) + L(\gamma)$, we see that

$$\rho(z_0, z_2) \leqslant \rho(z_0, z_1) + \rho(z_1, z_2) \ .$$

$\square$

Proposition 4.4   Isometries of $\mathbb{D}$.
Each symmetry in $\mathrm{Aut}(\mathbb{D})$ is an isometry for the hyperbolic metric.

Proof:

Let $T : \mathbb{D} \to \mathbb{D}$ be a complex differentiable map. An infinitesimal displacement from w to $w + dw$ is mapped by T to the displacement from $T(w)$ to $T(w + dw)$ and $T(w + dw)$ is equal to $T(w) + T'(w)dw$ to the first order in dw. Hence, T will be an isometry provided that

$$\frac{2|T'(w)||dw|}{1 - |T(w)|^2} = \frac{2|dw|}{1 - |w|^2} \ .$$

More formally, let $\gamma : [a, b] \to \mathbb{D}$ be a smooth curve. Then

$$L(\gamma) = \int_a^b \frac{2}{1 - |\gamma(t)|^2} \, |\gamma'(t)| \, dt$$

and

$$L(T \circ \gamma) = \int_a^b \frac{2}{1 - |(T \circ \gamma)(t)|^2} |(T \circ \gamma)'(t)| \, dt = \int_a^b \frac{2}{1 - |(T \circ \gamma)(t)|^2} |T'(\gamma(t))||\gamma'(t)| \, dt \ .$$

So the two lengths will be equal provided that

$$\frac{2|T'(w)|}{1 - |T(w)|^2} = \frac{2}{1 - |w|^2}$$

for each point $w = \gamma(t)$ on the curve γ.

For the map $T : z \mapsto \dfrac{az + b}{\bar{b}z + \bar{a}}$ with $|a|^2 - |b|^2 = 1$ we have

$$\frac{2|T'(w)|}{1 - |T(w)|^2} = \frac{2|1/(\bar{b}w + \bar{a})^2|}{1 - |(aw + b)/(\bar{b}w + \bar{a})|^2} = \frac{2}{|\bar{b}w + \bar{a}|^2 - |aw + b|^2} = \frac{2}{1 - |w|^2} \ .$$

So each $T \in \mathrm{Aut}^+(\mathbb{D})$ preserves the hyperbolic length of curves.

It is clear that conjugation $z \mapsto \bar{z}$ also preserves the hyperbolic length of curves. Therefore each map $\overline{T} : z \mapsto \overline{\left(\dfrac{az + b}{\bar{b}z + \bar{a}} \right)}$ in $\mathrm{Aut}^-(\mathbb{D})$ also does so.

Since each $T \in \mathrm{Aut}(\mathbb{D})$ preserves the lengths of all curves in $\mathbb{D}$, we must have $\rho(T(z_0), T(z_1)) = \rho(z_0, z_1)$. So T preserves the hyperbolic distance between points. $\square$

(In fact, every isometry of the hyperbolic plane $\mathbb{D}$ is one of the maps in $\mathrm{Aut}(\mathbb{D})$. Try to prove this.)

4.2 Hyperbolic geodesics

A *hyperbolic geodesic from z_0 to z_1 in $\mathbb{D}$* is a smooth curve in $\mathbb{D}$ from z_0 to z_1 that has minimal length. This means that the length of the geodesic is equal to the hyperbolic distance $\rho(z_0, z_1)$ from z_0 to z_1. It is not immediately clear that there is a curve that attains the length $\rho(z_0, z_1)$. However, we will show that there is and that it traces out a segment of the hyperbolic line through z_0 and z_1.

As a first example, let us find the geodesic from 0 to a point $w = Re^{i\alpha} \in \mathbb{D}$. Suppose that $\gamma : [0, 1] \to \mathbb{D}$ is a smooth path from 0 to w, say $\gamma(t) = r(t)e^{i\theta(t)}$. Then set $\beta(t) = r(t)e^{i\alpha}$. The hyperbolic length of γ is

$$L(\gamma) = \int_\gamma ds = \int_0^1 \frac{2|\gamma'(t)|}{1 - |\gamma(t)|^2} \, dt = \int_0^1 \frac{2|r'(t) + i\theta'(t)r(t)|}{1 - r(t)^2} \, dt \ .$$

while the hyperbolic length of β is

$$L(\beta) = \int_0^1 \frac{2|r'(t)|}{1 - r(t)^2} \, dt \ .$$

Hence we must have $L(\beta) \leqslant L(\gamma)$. Moreover, the inequality is strict unless $\theta(t)$ is constant. Thus the geodesic from 0 to w must trace out the radial line segment from 0 to w. We calculated the length of this path as $\log \dfrac{1 + |w|}{1 - |w|}$. Therefore $\rho(0, w) = \log \dfrac{1 + |w|}{1 - |w|}$.

Proposition 4.5   Hyperbolic geodesics
For any two distinct points z_0, z_1 in $\mathbb{D}$, the geodesic from z_0 to z_1 traces out the segment of the hyperbolic line through z_0 and z_1 that lies between the two points.

Proof:

We have seen that the curve from 0 to w with shortest hyperbolic length is a radial path. This traces out part of the diameter of $\mathbb{D}$ that passes through 0 and w.

Now suppose that z_0, z_1 are two distinct points of $\mathbb{D}$. Then there is an isometry

$$T : z \mapsto \frac{z + z_0}{1 + \overline{z_o}z}$$

that maps 0 to z_0 and the point $w = \frac{z_1 - z_0}{1 - \overline{z_0}z_1}$ to z_1. This must map the shortest hyperbolic path from 0 to w to the shortest hyperbolic path from z_0 to z_1. Since T maps hyperbolic lines to hyperbolic lines, the proof is complete. □

The argument in this proof also shows that $\rho(z_0, z_1) = \rho(T(0), T(w)) = \rho(0, w)$. Therefore,

$$\rho(z_0, z_1) = \rho(0, w) = \log \frac{1 + \left| \frac{z_1 - z_0}{1 - \overline{z_0}z_1} \right|}{1 - \left| \frac{z_1 - z_0}{1 - \overline{z_0}z_1} \right|} \ .$$

This is too complicated to be useful but, we can obtain more usable expressions for certain functions of $\rho(z_0, z_1)$. Write ρ for the hyperbolic distance $\rho(z_0, z_1)$. Then

$$\rho = \log(1 + R)/(1 - R) \qquad \text{for } R = \left| \frac{z_1 - z_0}{1 - \overline{z_0}z_1} \right| \ .$$

Inverting this gives $R = \tanh \frac{1}{2}\rho$. Therefore,

$$\sinh \tfrac{1}{2}\rho = \frac{R}{\sqrt{1 - R^2}} = \frac{|z_1 - z_0|}{\sqrt{1 - |z_0|^2}\sqrt{1 - |z_1|^2}} \ .$$

(This is the analogue of the formula for spherical geometry in Proposition 3.1.) Similarly,

$$\sinh \rho = \frac{2R}{1 - R^2} \qquad \text{and} \qquad \cosh \rho = \frac{1 + R^2}{1 - R^2} \ . \tag{†}$$

These formulae will enable us to prove the hyperbolic sine and cosine rules.

4.3 Hyperbolic triangles

A *hyperbolic triangle* Δ consists of three vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{D}$ and three geodesics from $\boldsymbol{B}$ to $\boldsymbol{C}$, from $\boldsymbol{C}$ to $\boldsymbol{A}$, and from $\boldsymbol{A}$ to $\boldsymbol{B}$. The geodesics are the sides of Δ and their lengths will be denoted by a, b, and c respectively. The angles at the vertices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are α, β and γ respectively.

Proposition 4.6   Hyperbolic cosine rule I
For the hyperbolic triangle Δ

$$\cosh a = \cosh b \cosh c - \sinh b \sinh c \cos \alpha \ .$$

Proof:

We may apply an isometry of $\mathbb{D}$ to ensure that $\boldsymbol{A} = 0$. Then, by rotating about 0 (which is also a hyperbolic isometry), we can ensure that B lies on the positive real axis. Then we have

$$\boldsymbol{A} = 0 \quad \boldsymbol{B} = |\boldsymbol{B}| \ , \quad \boldsymbol{C} = |\boldsymbol{C}|e^{i\alpha} \ .$$

Now we have

$$\cosh a = \cosh \rho(\boldsymbol{B}, \boldsymbol{C}) = \frac{1 + R^2}{1 - R^2} \qquad \text{where} \qquad R = \left| \frac{\boldsymbol{B} - \boldsymbol{C}}{1 - \overline{\boldsymbol{B}}\boldsymbol{C}} \right| \ .$$

Simplifying this yields

$$\begin{aligned}
\cosh a &= \frac{|1 - \overline{\boldsymbol{B}}\boldsymbol{C}|^2 + |\boldsymbol{B} - \boldsymbol{C}|^2}{|1 - \overline{\boldsymbol{B}}\boldsymbol{C}|^2 - |\boldsymbol{B} - \boldsymbol{C}|^2} = \frac{(1 + |\boldsymbol{B}|^2)(1 + |\boldsymbol{C}|^2) - 2(\overline{\boldsymbol{B}}\boldsymbol{C} + \boldsymbol{B}\overline{\boldsymbol{C}})}{(1 - |\boldsymbol{B}|^2)(1 - |\boldsymbol{C}|^2)} \\
&= \left(\frac{1 + |\boldsymbol{B}|^2}{1 - |\boldsymbol{B}|^2} \right) \left(\frac{1 + |\boldsymbol{C}|^2}{1 - |\boldsymbol{C}|^2} \right) - \left(\frac{2|\boldsymbol{B}|}{1 - |\boldsymbol{B}|^2} \right) \left(\frac{2|\boldsymbol{C}|}{1 - |\boldsymbol{C}|^2} \right) \left(\frac{e^{i\alpha} + e^{-i\alpha}}{2} \right) \\
&= \cosh b \cosh c - \sinh b \sinh c \cos \alpha
\end{aligned}$$

because of the formulae (†) above. $\qquad\qquad\square$

Proposition 4.7   Hyperbolic sine rule
For the hyperbolic triangle Δ

$$\frac{\sinh a}{\sin \alpha} = \frac{\sinh b}{\sin \beta} = \frac{\sinh c}{\sin \gamma} \ .$$

Proof:

Let $\boldsymbol{C}^*$ be the reflection of $\boldsymbol{C}$ in the hyperbolic line through $\boldsymbol{A}$ and $\boldsymbol{B}$. As in the proof of the cosine rule, we may assume that

$$\boldsymbol{A} = 0 \quad \boldsymbol{B} = |\boldsymbol{B}| \ , \quad \boldsymbol{C} = |\boldsymbol{C}|e^{i\alpha} \ .$$

Then $\boldsymbol{C}^* = |\boldsymbol{C}|e^{-i\alpha}$. Consequently, the hyperbolic distance from $\boldsymbol{C}$ to $\boldsymbol{C}^*$ satisfies

$$\sinh \tfrac{1}{2}\rho(\boldsymbol{C}, \boldsymbol{C}^*) = \frac{|\boldsymbol{C} - \boldsymbol{C}^*|}{\sqrt{1 - |\boldsymbol{C}|^2}\sqrt{1 - |\boldsymbol{C}^*|^2}} = \frac{2|\boldsymbol{C}| \sin \alpha}{1 - |\boldsymbol{C}|^2} \ .$$

Now (†) shows that $2|\boldsymbol{C}|/(1 - |\boldsymbol{C}|^2) = \sinh \rho(0, \boldsymbol{C}) = \sinh b$. So we see that

$$\sinh \tfrac{1}{2}\rho(\boldsymbol{C}, \boldsymbol{C}^*) = \sinh b \sin \alpha \ .$$

If we interchange $\boldsymbol{A}$ and $\boldsymbol{B}$ the point $\boldsymbol{C}^*$ is unchanged, so we must have

$$\sinh \tfrac{1}{2}\rho(\boldsymbol{C}, \boldsymbol{C}^*) = \sinh a \sin \beta \ .$$

Consequently,

$$\frac{\sinh a}{\sin \alpha} = \frac{\sinh b}{\sin \beta} \ .$$

By permuting the vertices we obtain the desired result. $\qquad\qquad\square$

There is also a second Cosine rule in which the rôles of angles and edge lengths are interchanged.

Proposition 4.8   Hyperbolic cosine rule II
For the hyperbolic triangle Δ

$$\cos\alpha = -\cos\beta\cos\gamma + \sin\beta\sin\gamma\cosh a \ .$$

Proof:
See Question 13 on the second Example Sheet. □

4.4 The upper half-plane model for the hyperbolic plane

The disc is only one model for the hyperbolic plane and it is sometimes convenient to use others. We can use any bijection $f : \mathbb{D} \to \Omega$ to transfer the hyperbolic metric from $\mathbb{D}$ to another region Ω. An important case is when we use the Möbius transformation

$$J : \mathbb{D} \to \mathcal{H}^+ \ ; \quad w \mapsto i\left(\frac{1+w}{1-w}\right)$$

that maps the disc $\mathbb{D}$ onto the *upper half-plane* $\mathcal{H}^+ = \{z = x + iy \in \mathbb{C} : y > 0\}$. This has inverse

$$J^{-1} : \mathcal{H}^+ \to \mathbb{D} \ ; \quad z \mapsto \frac{1+iz}{-1+iz} \ .$$

Since J is a Möbius transformation, hyperbolic lines in $\mathbb{D}$ are mapped to circles or straight lines in the upper half-plane. Since J is conformal, these circles cross the boundary $\partial\mathcal{H}^+ = \mathbb{R} \cup \{\infty\}$ orthogonally. Thus the hyperbolic lines in the upper half-plane are either vertical straight lines or semi-circles between points of $\mathbb{R}$.

To find the hyperbolic metric on the upper half-plane, observe that for $z = x + iy = J(w)$ we have

$$\frac{dw}{dz} = \frac{2i}{(-1+iz)^2} \qquad \text{so} \qquad |dw| = \frac{2}{|-1+iz|^2}|dz| \ .$$

Also,

$$1 - |w|^2 = \frac{|-1+iz|^2 - |1+iz|^2}{|-1+iz|^2} = \frac{4y}{|-1+iz|^2} \ .$$

These imply that

$$ds = \frac{2|dw|}{1-|w|^2} = \frac{|dz|}{y} \ .$$

This gives the hyperbolic metric on the upper half-plane.

Proposition 4.9   Isometries for the upper half-plane.
For real numbers a, b, c, d with $ad - bc = 1$ the Möbius transformation

$$T : z \mapsto \frac{az+b}{cz+d}$$

is an isometry of the upper half-plane with the hyperbolic metric.

Proof:

It is clear that T maps the extended real line $\mathbb{R} \cup \{\infty\}$ onto itself. So it must map $\mathcal{H}^+$ onto either the upper half-plane or the lower half-plane. Since

$$T(i) = \frac{ai+b}{ci+d} = \frac{(bd-ac) + i(ad-bc)}{|ci+d|^2} \in \mathcal{H}^+$$

we see that T maps $\mathcal{H}^+$ onto itself.

We need to check that

$$\frac{|T'(z)|}{\text{Im}(T(z))} = \frac{1}{\text{Im}(z)} \ .$$

Now $T'(z) = 1/(cz+d)^2$ and

$$\text{Im}(T(z)) = \frac{\left(\frac{az+b}{cz+d}\right) - \overline{\left(\frac{az+b}{cz+d}\right)}}{2i} = \frac{(az+b)(c\overline{z}+d) - (a\overline{z}+b)(cz+d)}{2i|cz+d|^2} = \frac{z - \overline{z}}{2i|cz+d|^2} \ .$$

So T is an isometry. $\qquad\square$

These are all the orientation preserving isometries of $\mathcal{H}^+$ and indeed all of the Möbius transformations that map $\mathcal{H}^+$ onto itself. For suppose that $T : z \mapsto \dfrac{az+b}{cz+d}$ is a Möbius transformation that maps the upper half-plane onto itself. Then it must map the boundary $\mathbb{R} \cup \{\infty\}$ onto itself. So the ratios $T(0) = b/d$, $T(\infty) = a/c$, $T^{-1}(0) = -b/a$ and $T^{-1}(\infty) = -d/c$ must all be real or ∞. By multiplying all of them by a scalar, we can ensure that all of them are real. Indeed, we can ensure that $ad-bc = \pm 1$. Now

$$\text{Im}(T(z)) = \frac{\text{Im}(z)}{|cz+d|^2} \ .$$

So T will only map the upper half-plane to the upper half-plane when $ad - bc = 1$. Otherwise it will map onto the lower half-plane.

We have shown that each matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{R})$$

gives an isometry of $\mathcal{H}^+$. Two matrices give the same Möbius transformation when one is plus or minus one times the other. So the group of orientation preserving isometries of $\mathcal{H}^+$ is $\text{SL}(2, \mathbb{R})/\{-I, I\}$. We call this quotient the *projective special linear group* $\text{PSL}(2, \mathbb{R})$.

We can use either the disc model or the upper half-plane model to establish results in hyperbolic geometry. Generally, we use whichever makes our calculations easier. Use the disc model when there is a special point in the hyperbolic plane (send it to 0) and the upper half-plane model when there is a special point on the boundary of the hyperbolic plane (send it to ∞) or a special hyperbolic line (send it to the imaginary axis).

4.5 The area of a hyperbolic triangle

So far we have only considered hyperbolic triangles with vertices in the hyperbolic plane. However, it is useful to allow some or all of the vertices to lie on the boundary. Since the hyperbolic lines cross the boundary orthogonally, the angle at such a vertex must be 0.

**Proposition 4.10** Area of a hyperbolic triangle.
A hyperbolic triangle with angles α, β and γ has hyperbolic area $\pi - (\alpha + \beta + \gamma)$. Hence the sum of the angles of a hyperbolic triangle is at most π.

Proof:

We will first prove this in the special case where two of the vertices of the triangle Δ lie on the boundary of hyperbolic space, so the corresponding angles are 0. We may use the upper half-plane model and assume that $\boldsymbol{C} = \infty$, $\boldsymbol{B} = 1$ and that $\boldsymbol{A} = e^{i\theta}$ with $0 \leqslant \theta \leqslant \pi$. Then the angle α at $\boldsymbol{A}$ is $\pi - \theta$. The hyperbolic area of this triangle is

$$\int_{\cos\theta}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2} \, dy \, dx = \int_{\cos\theta}^{1} \frac{1}{\sqrt{1-x^2}} \, dx = \theta \ .$$

So, for a triangle with one angle α and the other two angles 0, the hyperbolic area is $\pi - \alpha$.
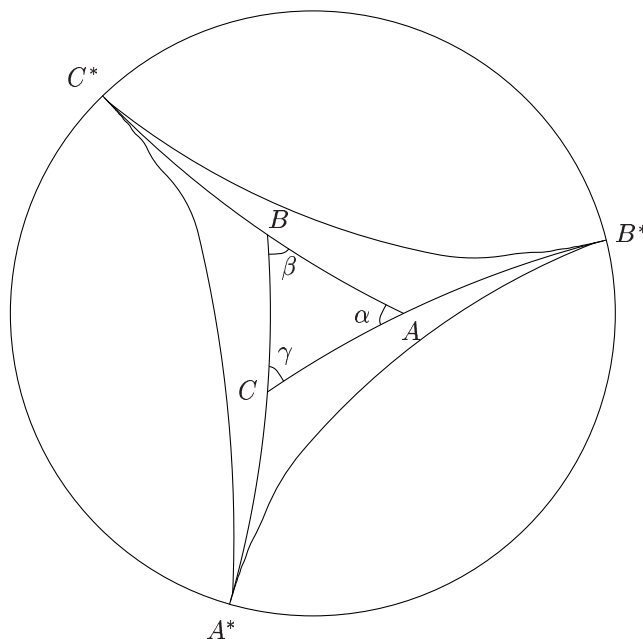
Now consider any hyperbolic triangle Δ. Extend the sides of the triangle as shown to meet the boundary of $\mathbb{D}$ at points $\boldsymbol{A}^*, \boldsymbol{B}^*$ and $\boldsymbol{C}^*$. Then our first result shows that

$$\text{Area}(\boldsymbol{A}^*\boldsymbol{B}^*\boldsymbol{C}^*) = \pi$$
$$\text{Area}(\boldsymbol{A}\boldsymbol{B}^*\boldsymbol{C}^*) = \alpha$$
$$\text{Area}(\boldsymbol{A}^*\boldsymbol{B}\boldsymbol{C}^*) = \beta$$
$$\text{Area}(\boldsymbol{A}^*\boldsymbol{B}^*\boldsymbol{C}) = \gamma \ .$$

So we see that

$$\begin{aligned} \text{Area}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) &= \text{Area}(\boldsymbol{A}^*\boldsymbol{B}^*\boldsymbol{C}^*) - \text{Area}(\boldsymbol{A}\boldsymbol{B}^*\boldsymbol{C}^*) - \text{Area}(\boldsymbol{A}^*\boldsymbol{B}\boldsymbol{C}^*) - \text{Area}(\boldsymbol{A}^*\boldsymbol{B}^*\boldsymbol{C}) \\ &= \pi - \alpha - \beta - \gamma \end{aligned}$$

as required. $\qquad \square$

For any angles α, β, γ with $\alpha + \beta + \gamma < \pi$ there is a hyperbolic triangle Δ with these angles, for we may use the cosine rule II to find the appropriate lengths of the sides. When

$$\alpha = \frac{\pi}{m} \; ; \quad \beta = \frac{\pi}{n} \; ; \quad \gamma = \frac{\pi}{p}$$

for some natural numbers m, n, p, with

$$S = \frac{1}{m} + \frac{1}{n} + \frac{1}{p}$$

strictly less than 1, reflections in the sides of Δ give a triangulation of the hyperbolic plane. There are infinitely many such triples. However, there are only finitely many triples with $S = 1$ or $S > 1$ and hence there are only finitely many corresponding triangulations of the Euclidean plane or the sphere.

4.6 *The hyperboloid model*

The sphere is the subset $\{\boldsymbol{x} : \boldsymbol{x} \cdot \boldsymbol{x} = 1\}$ of $\mathbb{R}^3$. In many contexts we use a different inner product from the dot product $\boldsymbol{x} \cdot \boldsymbol{y}$. In particular, we can study the indefinite inner product on $\mathbb{R}^3$ given by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = -x_0 y_0 + x_1 y_1 + x_2 y_2 \ .$$

This is <u>not</u> positive definite for $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = -x_0^2 + x_1^2 + x_2^2$ can be negative or positive. (This inner product is used in Special Relativity where we define the "proper length" of a 4-vector (t, x_1, x_2, x_3) to be the square root of $-c^2 t^2 + x_1^2 + x_2^2 + x_3^2$.) The subset corresponding to the unit sphere is

$$\{(x_0, x_1, x_2) \in \mathbb{R}^3 : -x_0^2 + x_1^2 + x_2^2 = -1\} \ .$$

This is a two-sheeted hyperboloid with one sheet

$$Q^+ = \{(x_0, x_1, x_2) \in \mathbb{R}^3 : -x_0^2 + x_1^2 + x_2^2 = -1 \text{ and } x_0 > 0\}$$

lying inside the *forward light cone*: $\{\boldsymbol{x} : -x_0^2 + x_1^2 + x_2^2 \geqslant 0 \text{ and } x_0 \geqslant 0\}$ and the other lying in the backward light cone. We can think of the hyperboloid Q^+ as giving us another model for hyperbolic geometry.

It is convenient to project stereographically from the South pole $(-1, 0, 0)$ rather than the North pole. For $\boldsymbol{x} \in Q^+$ the straight line from $\boldsymbol{x}$ to $(-1, 0, 0)$ crosses the plane $\{x_0 = 0\}$ at a point $(0, u, v)$. We write $\pi(\boldsymbol{x})$ for the complex number $w = u + iv$. A simple calculation gives

$$\pi(\boldsymbol{x}) = \frac{x_1 + i x_2}{1 + x_0} \ .$$

Since $-x_0^2 + x_1^2 + x_2^2 = -1$, we have

$$1 - |w|^2 = \frac{(1 + x_0)^2 - x_1^2 - x_2^2}{(1 + x_0)^2} = \frac{2 + 2x_0}{(1 + x_0)^2} = \frac{2}{1 + x_0} \ .$$

So $w = \pi(\boldsymbol{x})$ lies in the unit disc $\mathbb{D} = \{w \in \mathbb{C} : |w| < 1\}$.

On Q^+ an infinitesimal displacement $d\boldsymbol{x}$ has length ds with

$$ds^2 = \langle d\boldsymbol{x}, d\boldsymbol{x} \rangle = -dx_0^2 + dx_1^2 + dx_2^2 \ .$$

We want to write this in terms of the dw. First note that

$$dw = \frac{-(x_1 + i x_2)}{(1 + x_0)^2} dx_0 + \frac{dx_1 + i \, dx_2}{1 + x_0} \ .$$

32

So we obtain
$$|dw|^2 = \frac{x_1^2 + x_2^2}{(1 + x_0)^4} dx_0^2 - 2\frac{(x_1 dx_1 + x_2 dx_2) dx_0}{(1 + x_0)^3} + \frac{1}{(1 + x_0)^2}(dx_1^2 + dx_2^2) \ .$$
Since $-x_0^2 + x_1^2 + x_2^2 = -1$, we obtain $x_0 dx_0 = x_1 dx_1 + x_2 dx_2$. So the above formula simplifies to
$$|dw|^2 = \frac{x_0^2 - 1}{(1 + x_0)^4} dx_0^2 - 2\frac{x_0}{(1 + x_0)^3} dx_0^2 + \frac{1}{(1 + x_0)^2}(dx_1^2 + dx_2^2) = \frac{-dx_0^2 + dx_1^2 + dx_2^2}{(1 + x_0)^2} \ .$$
Therefore
$$ds^2 = \left(\frac{2|dw|}{1 - |w|^2}\right)^2 \ .$$
This is the hyperbolic metric on the disc $\mathbb{D}$, so we have shown that stereographic projection gives an isometry from Q^+ with the Riemannian metric $ds^2 = \langle d\boldsymbol{x}, d\boldsymbol{x} \rangle$ to the unit disc $\mathbb{D}$ with its hyperbolic metric. Thus Q^+ is another model for the hyperbolic plane.
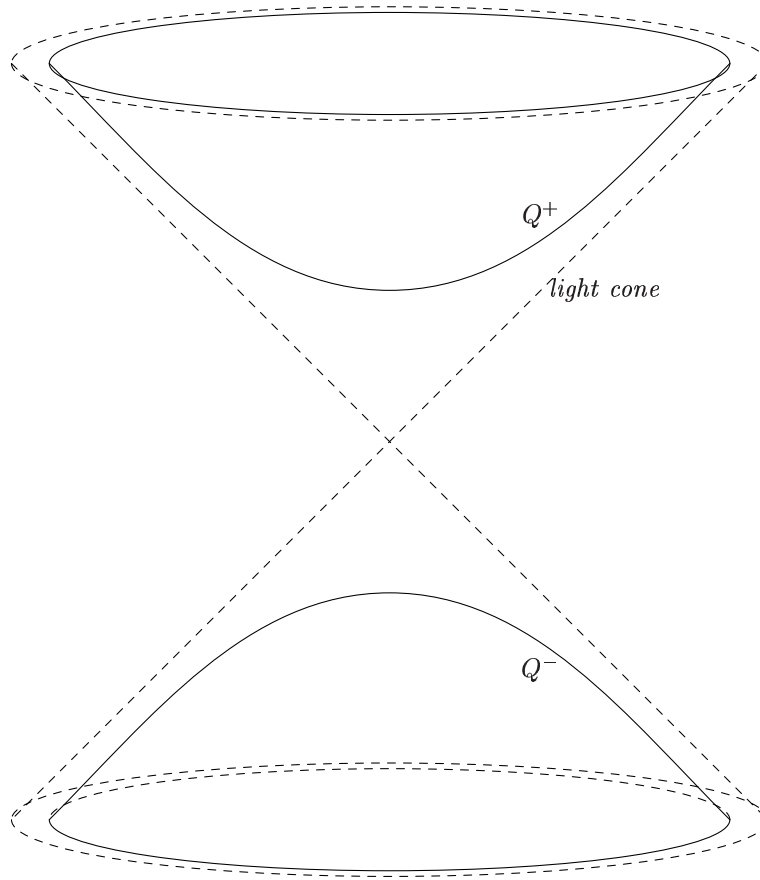
In many ways this hyperboloid model is algebraically simpler to work with than either the disc or the upper half-plane models. Indeed, we can use it in very much the same way as we used the sphere S^2 in section 2. For example, the hyperbolic lines in Q^+ are the intersections of Q^+ with 2-dimensional vector subspaces of $\mathbb{R}^3$. The inner product satisfies
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \cosh \rho(\boldsymbol{x}, \boldsymbol{y})$$
so
$$\sinh \tfrac{1}{2}\rho(\boldsymbol{x}, \boldsymbol{y}) = \tfrac{1}{2}\sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle} \ .$$
(You should compare this to the formula for the sphere given in Proposition 2.1:
$$\boldsymbol{x} \cdot \boldsymbol{y} = \cos d(\boldsymbol{x}, \boldsymbol{y})$$
which implies that
$$\sin \tfrac{1}{2}d(\boldsymbol{x}, \boldsymbol{y}) = \tfrac{1}{2}\sqrt{(\boldsymbol{x} - \boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})} = \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{y}|| \ .)$$

5. FINITE SYMMETRY GROUPS — THE PLATONIC SOLIDS

We will study finite groups of isometries of the Euclidean plane and 3-space.

5.1 Finite Isometry Groups

Proposition 5.1   Finite Groups of Isometries for $\mathbb{E}^N$.
If G is a finite group of isometries of $\mathbb{E}^N$, then there is a point $\boldsymbol{P} \in \mathbb{E}^N$ that is fixed by each isometry in G. Hence G is conjugate to a subgroup of $\mathrm{O}(N)$.

Proof:

First note that any isometry $T : \boldsymbol{x} \mapsto R\boldsymbol{x} + \boldsymbol{t}$ satisfies

$$T\left(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k\right) = \frac{1}{K}\sum_{k=1}^{K}T(\boldsymbol{x}_k) \ .$$

Choose any point $\boldsymbol{A} \in \mathbb{E}^N$ and set $\boldsymbol{P} = \dfrac{1}{|G|}\displaystyle\sum_{g \in G} g(\boldsymbol{A})$. For $T \in G$ we then have

$$T(\boldsymbol{P}) = T\left(\frac{1}{|G|}\sum_{g \in G} g(\boldsymbol{A})\right) = \frac{1}{|G|}\sum_{g \in G} Tg(\boldsymbol{A}) = \frac{1}{|G|}\sum_{h \in G} h(\boldsymbol{A}) \ .$$

So each $T \in G$ fixes $\boldsymbol{P}$.

Let U be the translation $\boldsymbol{x} \mapsto \boldsymbol{x} - \boldsymbol{P}$. Then each element of UGU^{-1} fixes the origin. So UGU^{-1} is a subgroup of $\mathrm{O}(N)$. $\qquad\square$

(Alternatively, we may argue as follows. Choose any point $\boldsymbol{A} \in \mathbb{E}^N$ and consider the finite orbit $\Omega = \{T(\boldsymbol{A}) : T \in G\}$ of $\boldsymbol{A}$. Let B be the closed Euclidean ball $\{\boldsymbol{P} \in \mathbb{E}^N : d(\boldsymbol{C}, \boldsymbol{P}) \leqslant R\}$ of smallest radius R that contains Ω. This ball is unique, for two balls of the same radius R intersect in a lune that lies within a ball of strictly smaller radius. However, each $T \in G$ maps the set Ω into itself and maps B to a closed ball of radius R that also contains Ω. Therefore $T(B) = B$. This certainly implies that T maps the centre of B to itself.)

Let K be a finite subgroup of rotations of $\mathbb{E}^2$. By conjugating this we may assume that each element of K fixes the origin. So $K \leqslant \mathrm{SO}(2)$.

Lemma 5.2   Finite Groups of rotations of the Euclidean plane.
Let K be a group of M rotations of $\mathbb{E}^2$ about $\boldsymbol{0}$. Then K is the cyclic group consisting of rotations through angles $2\pi k/M$ for $k = 0, 1, 2, \ldots, M - 1$.

Proof:

Choose the rotation $R \in K \setminus \{I\}$ that rotates through the smallest positive angle, say θ. If $S \in K$ is a rotation through an angle ϕ then we can find an integer k with $k\theta \leqslant \phi < (k+1)\theta$. So $R^{-k}S$ is a rotation in K through an angle $\phi - k\theta$ with $0 \leqslant \phi - k\theta < \theta$. The choice of R shows that $R^{-k}S$ must be the identity. Hence G is a cyclic group generated by R.

Each rotation in K has order a factor of M. There are only M such rotations so K must consist of I and all the rotations by $2\pi k/M$ for $k = 1, 2, \ldots, M - 1$. (R is the rotation through $2\pi/M$.) $\qquad\square$

When we come to consider groups that contain orientation reversing isometries we need to look at reflections more carefully.

Lemma 5.3 The Composition of Reflections.
Let U, V be reflections of $\mathbb{E}^2$ in the lines ℓ, m respectively. If ℓ and m are parallel, then VU is translation perpendicular to ℓ and m through twice the distance from ℓ to m. If ℓ and m meet at a point $\boldsymbol{P}$, then VU is a rotation about $\boldsymbol{P}$ through twice the angle from ℓ to m.

Proof:
    If ℓ and m are parallel, we may choose co-ordinates so that $\ell = \{(0, y) : y \in \mathbb{R}\}$ and $m = \{(d, y) : y \in \mathbb{R}\}$. Then

$$U : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -x \\ y \end{pmatrix} \quad \text{and} \quad V : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2d - x \\ y \end{pmatrix}$$

so we find that

$$VU : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2d + x \\ y \end{pmatrix} \ .$$

Thus VU is translation by a distance $2d$ perpendicular to ℓ and m.

    If ℓ and m intersect, we may choose co-ordinates so that $\ell = \{(x, 0) : x \in \mathbb{R}\}$ and $m = \{\lambda(\cos\theta, \sin\theta) : \lambda \in \mathbb{R}\}$. Then

$$U : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad V : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

so we find that

$$VU : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \ .$$

Thus VU is rotation through 2θ about the point $\boldsymbol{0}$ where ℓ and m meet. $\square$

    Now suppose that U is a reflection in a line ℓ and R a rotation through an angle ϕ about a point $\boldsymbol{P}$ on ℓ. Then the composite RU is also a reflection. For if V is the reflection in the line through $\boldsymbol{P}$ at an angle $\frac{1}{2}\phi$ to ℓ, then VU is equal to the rotation R. Hence $RU = V$.

Proposition 5.4 Finite Groups of Isometries of the Euclidean Plane.
Let G be a finite group of isometries of the Euclidean plane containing at least 3 elements. Then there is a regular polygon Π with M sides $(M = 2, 3, 4, \ldots)$ so G is either the dihedral group of all symmetries of Π or else the cyclic group of all rotational symmetries of Π.

It is easy to deal with cases where G has less than 3 elements. If G has only 1 element, it is the identity. If it has 2 elements, I and R, then R is either a reflection or a rotation through angle π.

Proof:
    By Proposition 5.1 we can assume that G is a subgroup of $O(2)$. Let K be the subgroup of G consisting of orientation preserving transformations. Then Lemma 5.2 shows that K is cyclic of order M and is generated by the rotation R through an angle $2\pi/M$. If G contains no orientation reversing transformations, then $G = K$ and it is clear that G is the rotational symmetry group of a regular M-gon.

    Suppose that G contain an orientation reversing transformation U. Then U is reflection in some line ℓ through the origin. So $R^k U$ is a reflection in the line ℓ_k obtained by rotating ℓ about $\boldsymbol{0}$ through an angle $\pi k/M$ $(k = 0, 1, 2, \ldots, M - 1)$. Conversely, if V is any orientation reversing isometry in G, then VU is orientation preserving, so $VU = R^k$ for some $k = 0, 1, 2, \ldots, M - 1$. Therefore $V = R^k U$.

    Choose a point $\boldsymbol{A}_0$ on ℓ but distinct from the origin and set $\boldsymbol{A}_k = R^k(\boldsymbol{A}_0)$. Then the vertices $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{M-1}$ are the vertices of a regular M-gon Π and G is its symmetry group. Note that G is dihedral: it contains the identity, $M - 1$ rotations $R, R^2, R^3, \ldots, R^{M-1}$ and M reflections $U, RU, R^2, \ldots, R^{M-1}U$. $\square$

We can also study groups starting with the reflections. Suppose that U, V are two reflections in lines ℓ, m that meet at an angle ϕ. Then $R = VU$ is a rotation through 2ϕ. If ϕ is not a rational multiple of π, then R is of infinite order and so the group generated by U and V must be infinite. However, if $\phi = k\pi/M$ for co-prime natural numbers k, M, then U, V generate the dihedral group:

$$G = D_{2M} = \{I, R, R^2, R^3, \ldots, R^{M-1}; U, RU, R^2U, \ldots, R^{M-1}U\}$$

of order $2M$. This has M reflections $U, RU = V, R^2U, \ldots, R^{M-1}U$ in lines $\ell_0, \ell_1, \ell_2, \ldots, \ell_{M-1}$. These lines divide $\mathbb{E}^2$ into $2M$ sectors, each of angle π/M. We say that the sectors *tessellate* $\mathbb{E}^2$. Every element of G maps this tessellation onto itself. Colour the sectors alternately black and white. Then the rotations in G map each sector onto a sector of the same colour, and reflections in G map each sector onto a sector of a different colour.

$M = 5$
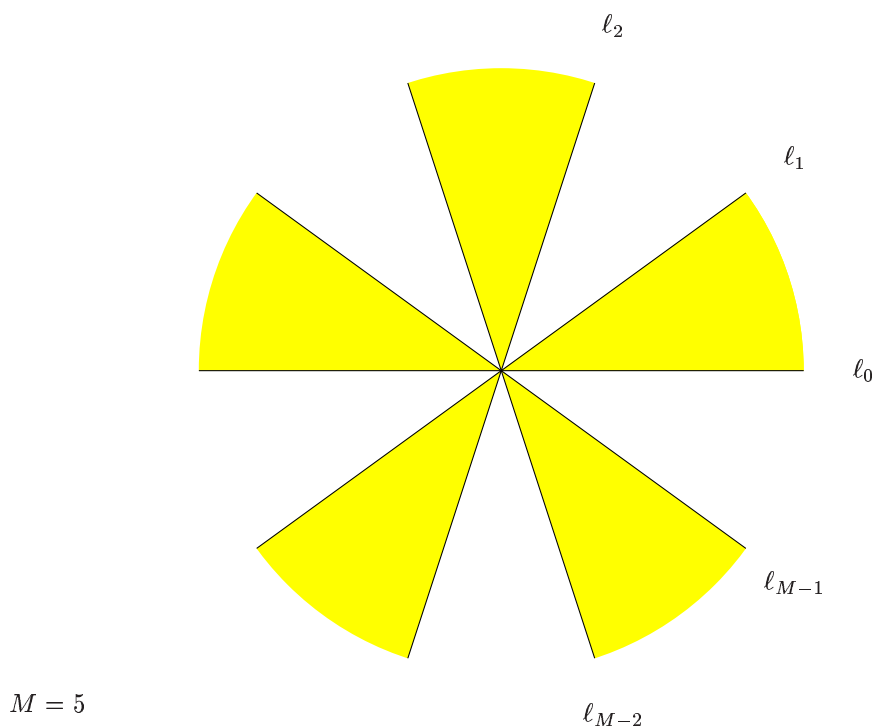
These results generalise to higher dimensions. Consider first Proposition 5.3. Let U, V be reflections in planes π_U, π_V in $\mathbb{E}^3$. If π_U and π_V intersect in a line n, then both U and V map all planes perpendicular to n into themselves. So Proposition 5.3 applies within each such plane and shows that VU is a rotation about the axis n through twice the angle from π_U to π_V. Similarly, if π_U and π_V are parallel, then VU is translation perpendicular to π_U and π_V through twice the distance from π_U to π_V.

We can also generalise the construction of the dihedral group from reflections. Suppose that $U_1, U_2, \ldots, U_K$ are reflections in hyperplanes $\pi_1, \pi_2, \ldots, \pi_K$ of $\mathbb{E}^N$ and that the angles between any two of these hyperplanes are π/m for some natural number $m \geqslant 2$. Then $U_1, U_2, \ldots, U_K$ generate a finite group G of isometries. The hyperplanes for all the reflections in G divide $\mathbb{E}^N$ into regions and G permutes these regions. Thus we obtain a tessellation of $\mathbb{E}^N$. This is significantly harder to prove and the details are omitted.

5.2 The Platonic Solids

The Platonic solids are the most regular polyhedra in $\mathbb{E}^3$. They are convex with each face being a regular Euclidean p-gon and q faces meeting at each vertex. The interior angles of a regular p-gon are $\pi - \dfrac{2\pi}{p}$. If q of these are to meet at each vertex, then $q\left(\pi - \dfrac{2\pi}{p}\right)$ must be strictly smaller than 2π. So

$$\frac{1}{2} < \frac{1}{p} + \frac{1}{q} \qquad \text{with} \qquad p, q \geqslant 3 \ .$$

The only solutions of this are $(p, q) = (3, 3), (4, 3), (3, 4), (5, 3), (3, 5)$ which give the regular tetrahedron, cube, octahedron, dodecahedron and icosahedron respectively. These are familiar objects.

It is often useful to think of these polyhedra as being on the sphere. Each of them can be embedded inside the unit sphere S^2 with each of the vertices lying on the sphere. Projecting radially out from the centre of the sphere maps each edge to an arc of a spherical line and each face to a regular spherical p-gon. These p-gons *tessellate the sphere*: they partition the surface into disjoint pieces that cover the entire sphere. Thus we can think of the Platonic solids as giving regular tessellations of the unit sphere S^2. We can compute the spherical area of each polygonal face and hence recover the result that there are only 5 Platonic solids. Later we will want to look at the symmetry groups of these polyhedra, so it is more convenient to first divide each face into triangles as in the result below.

Proposition 5.5 Platonic Solids
The Platonic solids have F faces, each of which is a regular p-gon and q faces meet at each vertex. The only five possibilities for F, p and q are:

	F	p	q
tetrahedron	4	3	3
cube	6	4	3
octahedron	8	3	4
dodecahedron	12	5	3
icosahedron	20	3	5

Proof:
Consider each face. This is a regular p-gon on the sphere S^2. Divide it into $2p$ triangles by joining the centre of the face to each of the p vertices and to each of the p midpoints of edges.

●	angles $\pi/2$
○	angles π/k
✕	angles π/m
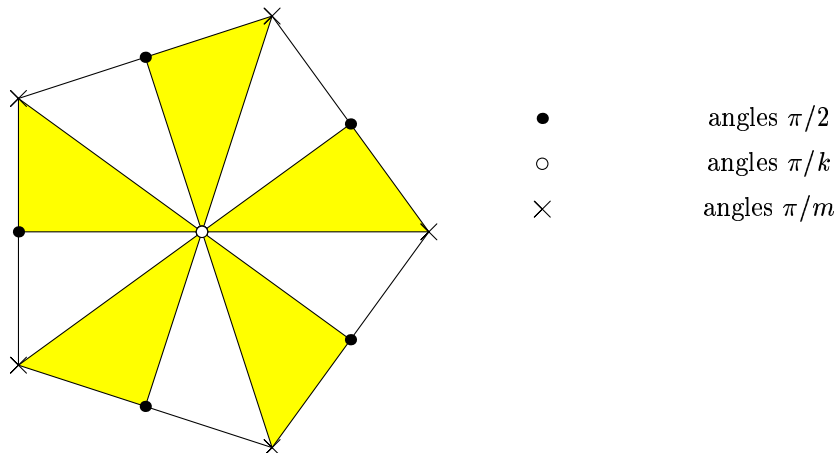
When we divide each face in this way, we obtain $N = 2Fp$ triangles that are all isometric. At the centre of each face $2p$ triangles meet so each has angle π/p. At each vertex of a face $2q$ triangles meet so each

has angle π/q. At the midpoint of each edge 4 triangles meet so each has angle $\pi/2$. Thus each triangle has area

$$\frac{\pi}{p} + \frac{\pi}{q} + \frac{\pi}{2} - \pi \ .$$

This must be the area of the sphere divided by N. Hence we must have

$$\frac{1}{p} + \frac{1}{q} - \frac{1}{2} = \frac{4}{N} \ .$$

The numbers p and q must be at least 3 and

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{2} + \frac{4}{N} > \frac{1}{2} \ .$$

Hence, neither k nor m can be more than 5. This gives the five possibilities described in the proposition.

	F	p	q	N
tetrahedron	4	3	3	24
cube	6	4	3	48
octahedron	8	3	4	48
dodecahedron	12	5	3	120
icosahedron	20	3	5	120

In each case, the N isometric triangles give a tessellation of the sphere. We pass from one triangle to a neighbour by reflecting in the common side. $\square$

These tessellations of the sphere by small triangles allow us to see what the group of isometries of the Platonic solids are. Let Π be one of the Platonic solids, drawn on the surface of the sphere. Each isometry T of the sphere that is a symmetry of Π must map one face of Π onto another. Hence it must map one of the small triangles in the tessellation onto another. Moreover, if T is any isometry of the sphere that does map one small triangle Δ in the tessellation onto another, then it must be a symmetry of the entire tessellation for we can obtain all of the triangles by reflecting repeatedly in the sides of Δ. The symmetry group of Π therefore permutes the triangles in the tessellation. Only the identity is in the stabilizer of Δ, so we see that the symmetry group of Π has order equal to the number of small triangles in the tessellation. We can also distinguish between the images of Δ under orientation preserving symmetries of Π and the images under orientation reversing symmetries. Suppose that we colour each image of Δ under an orientation preserving symmetry, as in the diagram above. Then half of the triangles are coloured. Every orientation preserving symmetry maps one coloured triangle to another while each orientation reversing symmetry maps one coloured triangle to an uncoloured one.

Only very special spherical triangles give a tessellation of the sphere when we repeatedly reflect in the sides. For suppose that Δ is a spherical triangle with angles α, β, γ at the vertices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. By reflecting in the two sides that meet at $\boldsymbol{A}$ we get a rotation about $\boldsymbol{A}$ through an angle 2α. If the successive images of Δ under these reflections are to fill out the region near the vertex $\boldsymbol{A}$ but not overlap, then this rotation must be of finite order. So $\alpha = \pi/p$ for some integer p. Similarly the other angles must be $\beta = \pi/q$, $\gamma = \pi/r$ for integers q, r. Since the angles of the triangle sum to more than π, we have

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$$

for integers $p, q, r \geqslant 2$. The only possibilities are that p, q, r are (in some order) $2, 2, m$ or $2, 3, 3$ or $2, 3, 4$ or $2, 3, 5$.

We have not yet shown that there are indeed regular polyhedra corresponding to each of the possibilities listed above. It is easy to describe the vertices for a tetrahedron, cube or octahedron explicitly and so to construct these. With a little more effort, you can do the same for the dodecahedron

and the icosahedron (see Example Sheet 1, question 14). Instead we will show how to construct the tessellation by small triangles from the proof of Proposition 5.5.

The tessellation with triangles we obtain in the five cases are not all different. Consider, for example, the cube and the octahedron. In each case we divide the sphere into 48 triangles each of which has angles $\pi/2, \pi/3$ and $\pi/4$. If we gather these triangles into groups of 8 about the vertices with angle $\pi/4$, we obtain the 6 faces of the cube. If we gather them into groups of 6 about the vertices with angle $\pi/3$, we obtain the 8 faces of the octahedron.

Sphere   Stereographic projection

——— edges of cube   ——— edges of octahedron
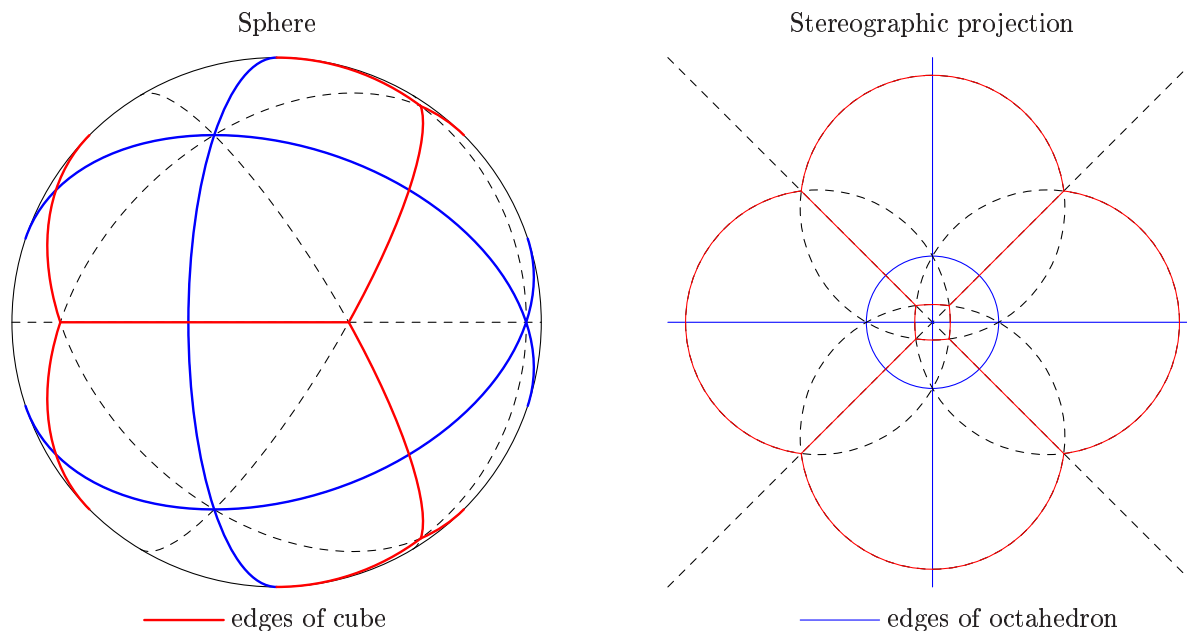
The vertices of the cube are the centre points of the faces of the octahedron and *vice versa*. We say that the cube and the octahedron are *dual* polyhedra. In a similar way, the tessellations for the dodecahedron and the icosahedron are the same, each consisting of 120 triangles with angles $\pi/2, \pi/3$ and $\pi/5$, so the dodecahedron and the icosahedron are dual. The tetrahedron is dual to itself.

5.3 Constructing tessellations of the sphere

Let (p, q) be one of the pairs $(3, 3), (4, 3), (3, 4), (5, 3), (3, 5)$. Then we can construct a spherical triangle Δ with angles $\pi/2, \pi/p$ and π/q at the vertices $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ respectively. We will show that reflecting in the sides of this triangle repeatedly gives a tessellation of the sphere. In this tessellation there are vertices where 4 of the images of $\boldsymbol{A}$ meet, vertices where $2p$ of the images of $\boldsymbol{B}$ meet and vertices where $2q$ of the images of C meet. We will call these *vertices of type $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$* respectively. The vertices of type $\boldsymbol{C}$ form the vertices of one of the Platonic solids and so we have a way to construct these solids.

Let us begin with the octahedron for which we can give the co-ordinates of the vertices explicitly. On the unit sphere the points

$$\boldsymbol{A} = (\tfrac{1}{2}\sqrt{2}, \tfrac{1}{2}\sqrt{2}, 0) \ , \quad \boldsymbol{B} = (\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}) \ , \quad \boldsymbol{C} = (1, 0, 0)$$

form the vertices of a spherical triangle Δ with angles $\pi/2, \pi/3$ and $\pi/4$. Reflecting in the sides gives a tessellation of the sphere using 24 triangles isometric to Δ. The vertices of type $\boldsymbol{C}$ are

$$(1, 0, 0) \ , \ (0, 1, 0) \ , \ (0, 0, 1) \ , \ (-1, 0, 0) \ , \ (0, -1, 0) \ , \ (0, 0, -1) \ ,$$

which form the vertices of a regular octahedron. The vertices of type $\boldsymbol{B}$ are

$$(\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}) \ , \ (\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}) \ , \ (\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}) \ , \ (\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}) \ ,$$
$$(-\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}) \ , \ (-\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}) \ , \ (-\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}) \ , \ (-\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}, -\tfrac{1}{3}\sqrt{3}) \ ,$$

which form the vertices of a cube. (The vertices of type $\boldsymbol{A}$ are not the vertices of a Platonic solid, instead they are the vertices of a cuboctahedron, which has 6 square faces and 8 equilateral triangular faces.)

This gives us the tessellation shown in the diagram below.
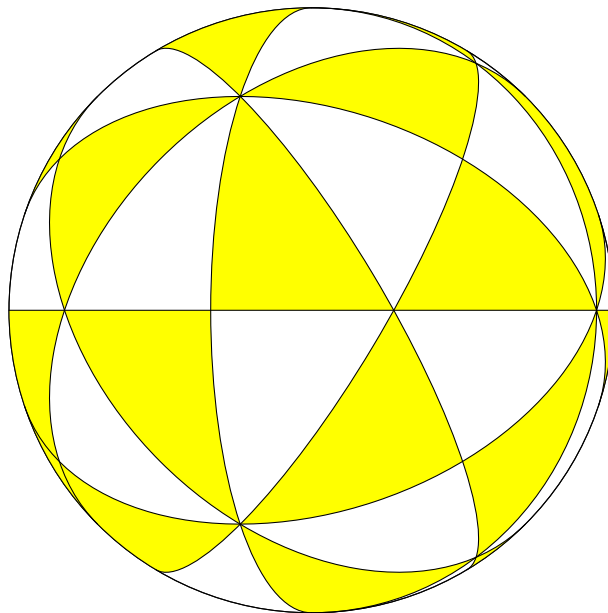
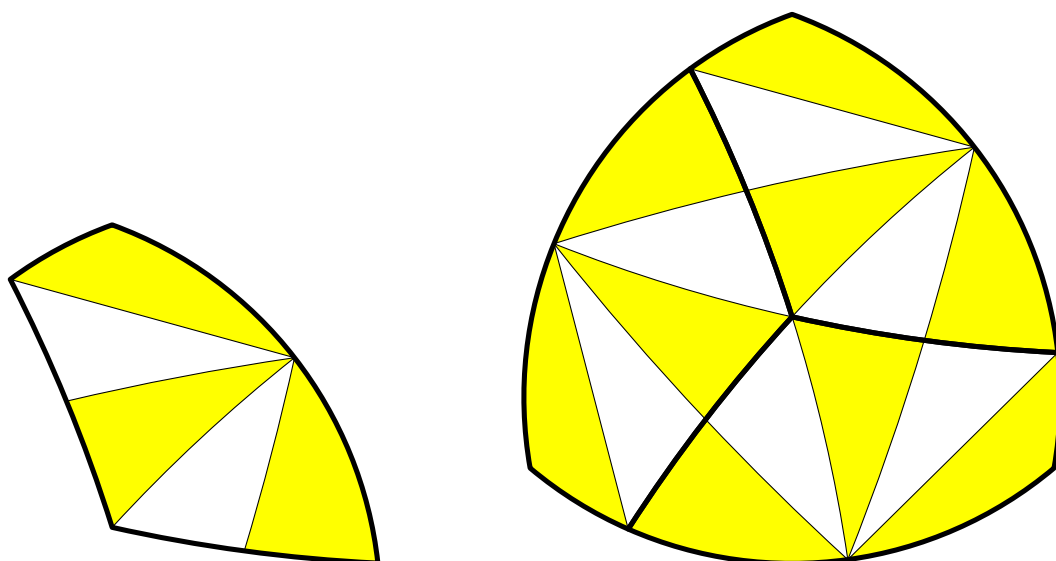Tessellation by $\pi/2, \pi/3, \pi/4$ spherical triangles

The symmetry group of this tessellation is identical to the symmetry group of the octahedron (and of the cube and cuboctahedron). Since the tessellation consists of 48 triangles, the group has order 48.

We can also construct the $\pi/2, \pi/3, \pi/3$ triangle we need for the tetrahedron. Indeed we can do this from the $(\pi/2, \pi/3, \pi/4)$-triangle given above. For consider the union of this triangle and its reflection in the side opposite the angle $\pi/3$. This gives a triangle Δ with angles $\pi/2, \pi/3, \pi/3$. Since 48 copies of the original triangle tessellate the sphere, we see that 24 copies of Δ also do so. Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be the vertices of Δ with the angles at these being $\pi/2, \pi/3$ and $\pi/3$ respectively. Then the vertices of type $\boldsymbol{C}$ form the vertices of a regular tetrahedron. The vertices of type $\boldsymbol{B}$ form the vertices of another regular tetrahedron. The vertices of type $\boldsymbol{A}$ form the vertices of a regular octahedron. (This corresponds to the observation that the midpoints of the edges of regular tetrahedron are the vertices of a regular octahedron. Alternatively, there are two tetrahedra whose vertices are also vertices of a cube; these tetrahedra intersect in an octahedron.)
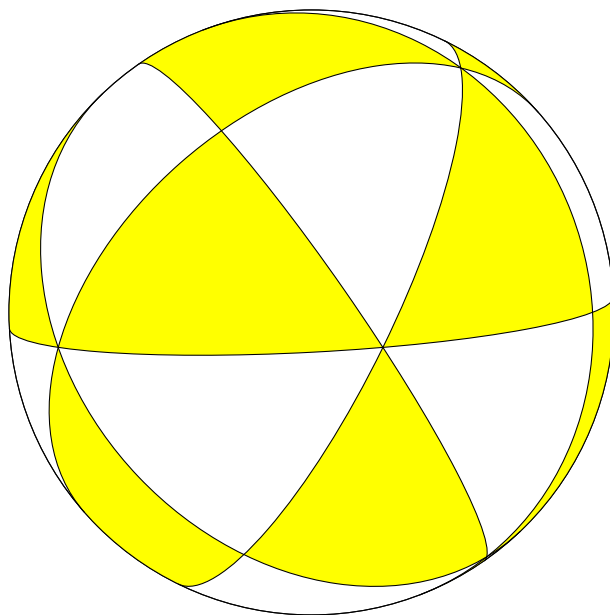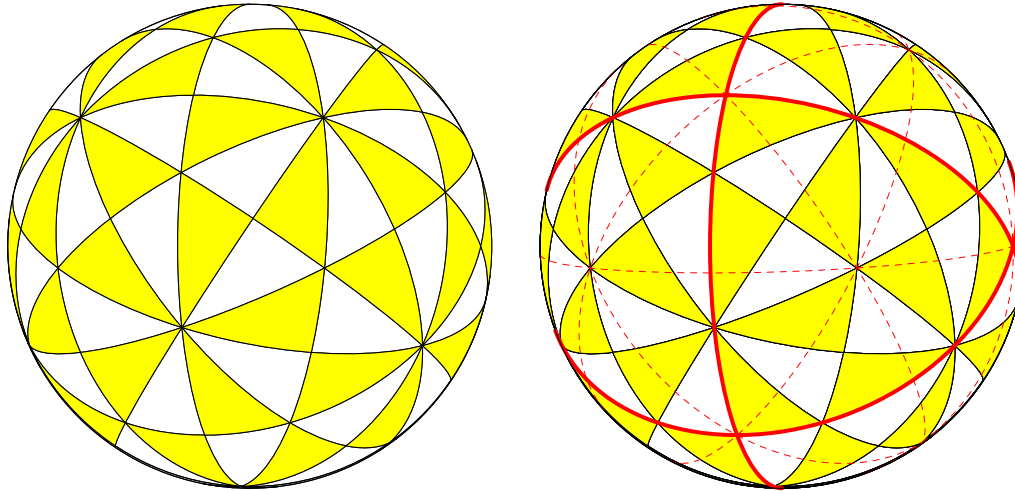

Tessellation by $\pi/2, \pi/3, \pi/3$ spherical triangles

Finally, we need to construct a tessellation of the sphere with $\pi/2, \pi/3, \pi/5$ triangles. We can certainly construct a spherical triangle Δ with these angles. By reflecting repeatedly in the sides that meet at angle $\pi/5$ we get the pattern of 5 triangles shown on the left below.

Three of these shapes fit together to form an equilateral triangle H with all angles $\frac{1}{2}\pi$, as shown on the right. This is one of the faces of a spherical octahedron, so we can fit 8 copies together to tessellate all of the sphere. This gives a tessellation of S^2 using $120 = 8 \times 3 \times 5$ triangles isometric to Δ. The diagram below shows the resulting tessellation of the sphere on the left with the tessellation for the octahedron superimposed on the right.

Tessellation by $\pi/2, \pi/3, \pi/5$ spherical triangles

We have now constructed the tessellations we required and hence have proved that the 5 Platonic solids really do exist.

5.4 Finite Symmetry Groups

Each of the Platonic solids has a finite symmetry group that is a subset of O(3). By using the orbit-stabiliser theorem from the Algebra and Geometry course, we can find all of the finite subgroups of O(3). For simplicity we will only consider the case where all of the isometries in the group preserve orientation. The general case then follows easily by considering the intersection of the group with SO(3).

Suppose that G is a finite group of orientation-preserving isometries of $\mathbb{E}^3$. By Proposition 5.1 we may assume that each element of G fixes the origin, so G is a subgroup of SO(3). Each isometry in G is either the identity or a rotation. Each rotation $R \in G$ fixes two points $\pm\boldsymbol{p}$ on the sphere S^2 that we call the *poles* of R. Let Ω be the set of all poles of rotations in G. If $\boldsymbol{p}$ is a a pole for R, then $T\boldsymbol{p}$ is a pole for the conjugate TRT^{-1}. Hence, the group G acts on the set Ω and permutes it. The orbit - stabiliser theorem will allow us to find all the possibilities for G.

Let $\Omega_1, \Omega_2, \ldots, \Omega_K$ be the orbits of G acting on Ω. For each pole $\boldsymbol{p} \in \Omega$ the stabiliser $\mathrm{Stab}(\boldsymbol{p})$ is a subgroup of G consisting of rotations about the axis through $\pm\boldsymbol{p}$. Lemma 5.2 now shows that $\mathrm{Stab}(\boldsymbol{p})$ is a cyclic group with finite order M_k. We can count the number of pairs $(T, \boldsymbol{p})$ with T a rotation in G and $\boldsymbol{p}$ a pole for T in two different ways. First, we can choose T as any of the $|G| - 1$ rotations in G and then choose $\boldsymbol{p}$ as either of the two poles of T. This gives $2(|G| - 1)$ pairs. Secondly, choose any $\boldsymbol{p} \in \Omega$. The rotations that have $\boldsymbol{p}$ as a pole are the non-identity elements in $\mathrm{Stab}(\boldsymbol{p})$. Hence the number of pairs is

$$\sum_{\boldsymbol{p} \in \Omega} (|\mathrm{Stab}(\boldsymbol{p})| - 1) = \sum_{\boldsymbol{p} \in \Omega} (M_k - 1) = \sum_{k=1}^{K} |\Omega_k|(M_k - 1) = \sum_{k=1}^{K} |G| \left(1 - \frac{1}{M_k}\right) .$$

Therefore, we have

$$2\left(1 - \frac{1}{|G|}\right) = \sum_{k=1}^{K}\left(1 - \frac{1}{M_k}\right) \ .$$

Now each M_k must be at least 2. So

$$2 > 2\left(1 - \frac{1}{|G|}\right) = \sum_{k=1}^{K}\left(1 - \frac{1}{M_k}\right) \geqslant \frac{K}{2}$$

and therefore K must be $1, 2$ or 3. If $K = 1$ we would need

$$1 + \frac{1}{M_1} = \frac{2}{|G|}$$

which is clearly impossible. For $K = 2$ we have $\Omega_1 = \{\boldsymbol{p}\}$ and $\Omega_2 = \{-\boldsymbol{p}\}$. So G is a cyclic group of rotations about the axis through $\pm\boldsymbol{p}$.

For $K = 3$ we can assume that $2 \leqslant M_1 \leqslant M_2 \leqslant M_3$. Then

$$1 < 1 + \frac{2}{|G|} = \frac{1}{M_1} + \frac{1}{M_2} + \frac{1}{M_3} \leqslant \frac{3}{M_1}$$

so we must have $M_2 = 2$. Then

$$\frac{1}{2} < \frac{1}{2} + \frac{2}{|G|} = \frac{1}{M_2} + \frac{1}{M_3} \leqslant \frac{2}{M_2}$$

so M_2 is 2 or 3. We will deal with these cases separately.

Suppose that $M_2 = 2$. Then M_3 can be any integer $M \geqslant 2$. Consequently

M_1	M_2	M_3	$|G|$	$|\Omega_1|$	$|\Omega_2|$	$|\Omega_3|$
2	2	M	$2M$	M	M	2

This shows that G is the rotational symmetry group of the pattern shown below. This is the tessellation of the sphere S^2 into M segments. (We can also think of G as the rotational symmetry group of a prism whose base is a regular M-sided polygon.)

● $= \Omega_1$

○ $= \Omega_2$

✕ $= \Omega_3$

43
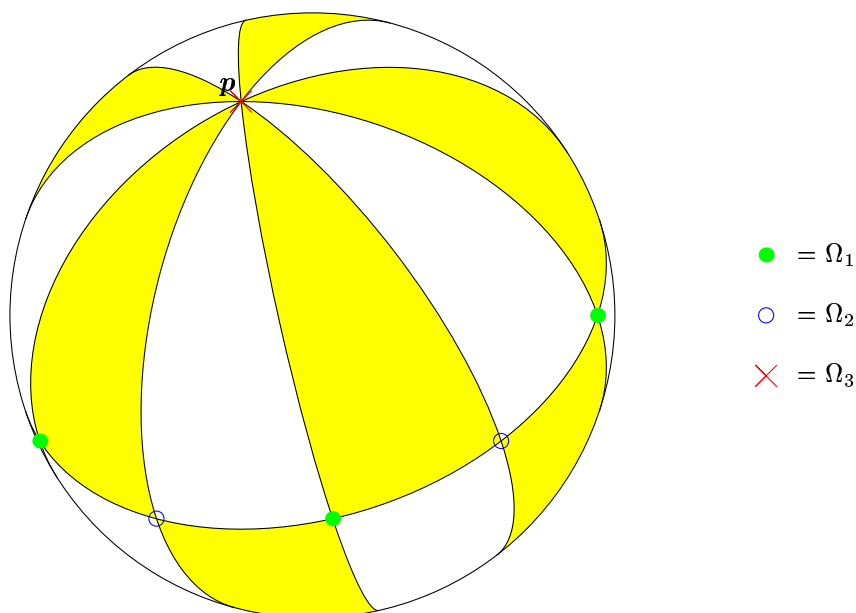
Now suppose that $M_2 = 3$. Then $\dfrac{1}{6} + \dfrac{2}{|G|} = \dfrac{1}{M_3}$, so $M_3 < 6$. There are three possibilities: (a) $M_3 = 3$; (b) $M_3 = 4$; and (c) $M_3 = 5$. In these cases we get:

	M_1	M_2	M_3	$|G|$	$|\Omega_1|$	$|\Omega_2|$	$|\Omega_3|$
(a)	2	3	3	12	6	4	4
(b)	2	3	4	24	12	8	6
(c)	2	3	5	60	30	20	12

These three groups are the rotational symmetry groups of the regular (a) tetrahedron, (b) octahedron, and (c) icosahedron. To see this, consider case (b) as an example. Here Ω_3 has 6 points. The stabiliser of each is a cyclic group of order $M_3 = 4$. Choose one pole $\boldsymbol{v} \in \Omega_3$. The stabiliser of $\boldsymbol{v}$ is a cyclic group of order 4; let R be a generator. Now $-\boldsymbol{v}$ is also a pole of R and has the same stabiliser. So it must be in Ω_3. There remain 4 other points in Ω_3 and these must be $\boldsymbol{w}, R(\boldsymbol{w}), R^2(\boldsymbol{w}), R^3(\boldsymbol{w})$ equally spaced around the equator separating $\boldsymbol{v}$ from $-\boldsymbol{v}$. Hence the points of Ω_3 are the 6 vertices of a regular octahedron.

Rather than thinking of the octahedron as having flat faces, think of it projected radially onto the sphere. Then each face is a spherical triangle with all angles $\pi/2$, and each edge is an arc of a great circle with length $\pi/2$. The points of Ω_2 are the centroids of the faces. The points of Ω_1 are the midpoints of the edges.
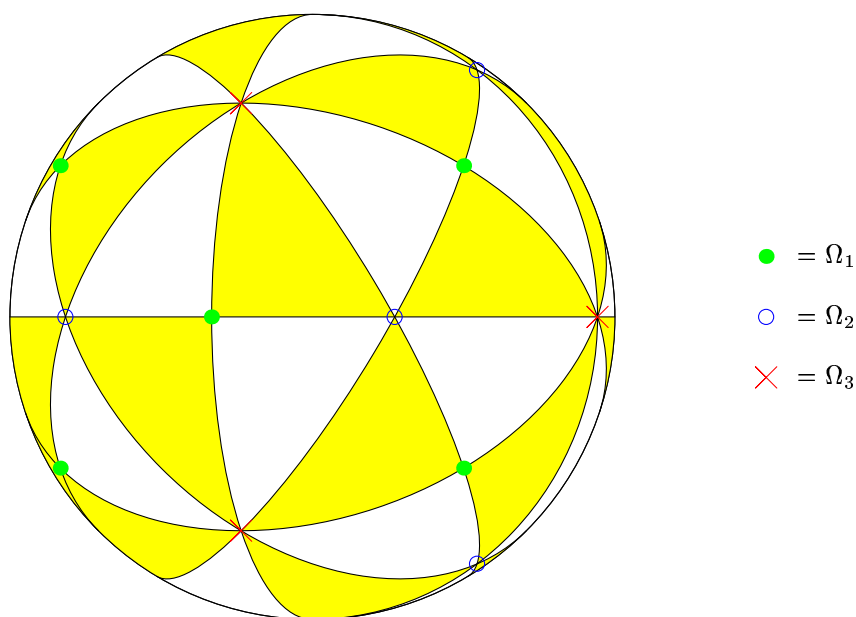
Note that the poles in Ω_2 are the vertices of a cube. This is the dual solid to the octahedron. The poles in Ω_1 are the vertices of a cuboctahedron (which has 6 square faces and 8 equilateral triangular faces). Each of these solids has the group G in case (b) as its rotational symmetry group.

6. *FINITE GROUPS GENERATED BY REFLECTIONS*

(This argument is not examinable for the Geometry course. It is only included to show that the results proven in section 6 for $\mathbb{E}^2$ generalise to higher dimensions.)

Let V be the sphere, Euclidean space or hyperbolic space of dimension m and let C be an open, convex subset of V bounded by a finite number of hyperplanes $H_1, H_2, \ldots, H_K$. Denote the two components of $V \setminus H_k$ by H_k^+ and H_k^-, with H_k^+ containing C. Then $C = \bigcap(H_k^+ : k = 1, 2, \ldots, K)$. We will assume that the angle (internal to C) between any two of the hyperplanes H_j and H_k is either 0 or π/m for some $m \in \{2, 3, 4, \ldots\}$.

Let R_k denote reflection in H_k and let G be the group of isometries of V generated by $R_1, R_2, \ldots, R_K$. Our aim is to prove that the cells $(gC : g \in G)$ tessellate V, so, for $g_1 \neq g_2$ in G the sets $g_1 C$ and $g_2 C$ are disjoint.

Each $g \in G$ can be written as $g = R_{i(1)} R_{i(2)} \ldots R_{i(L)}$ for some indices $i(r)$. The smallest value for L is the *length* of g and will be written $l(g)$. Consider the sequence of cells $C_0, C_1, C_2, \ldots, C_L$ given by

$$C_0 = C, \quad C_1 = R_{i(1)} C, \quad C_2 = R_{i(1)} R_{i(2)} C, \quad \ldots, \quad C_L = R_{i(1)} R_{i(2)} \ldots R_{i(L)} C = gC \ .$$

The cell C_n is obtained from C_{n-1} by reflecting in the side $R_{i(1)} R_{i(2)} \ldots R_{i(n-1)} H_{i(n)}$ of C_{n-1}. We will call such a sequence of cells a *chain from C to gC of length L*. Our first step is to see how to determine from the length function l whether gC is contained in one of the half-spaces $H_k^\pm$.

Proposition 6.1
For each $g \in G$ we have
$$gC \subset H_k^\pm \qquad \Leftrightarrow \qquad l(R_k g) = l(g) \pm 1 \ .$$

Moreover, either $gC \subset H_k^+$ or else $gC \subset H_k^-$.

Proof:

It is clear that $l(R_k g)$ and $l(g)$ differ by at most 1. Moreover, $l(g)$ is even if, and only if, g preserves orientation, so $l(R_k g)$ and $l(g)$ must have different parity. Thus $l(R_k g) = l(g) \pm 1$. It only remains to show that these conditions are equivalent to $gC \subset H_k^\pm$.

Consider the inductive hypothesis:

Q_L: For $g \in G$ with $l(g) \leqslant L$ we have $gC \subset H_k^\pm \quad \Leftrightarrow \quad l(R_k g) = l(g) \pm 1$.

The only element $g \in G$ with $L(g) = 0$ is the identity, so Q_0 is true. Assume that Q_{L-1} is true and that $g \in G$ has $l(g) = L$. Then $g = R_j g'$ for some j and some $g' \in G$ with $l(g') = L - 1$. We will consider separately the two cases $j = k$ and $j \neq k$.

If $j = k$ then $gC = R_k g' C$, so $gC \subset H_k^\pm$ is equivalent to $g'C \subset H_k^\mp$. Also $R_k g' = g$ so $l(R_k g) = l(g) \pm 1$ is equivalent to $l(R_k g') = l(g') \mp 1$. Therefore, the hypothesis Q_{L-1} applied to g' proves Q_L for g.

Now suppose that $j \neq k$. Let F be the subgroup of G generated by R_j and R_k. If the angle between H_j and H_k is π/m for some $m \in \{2, 3, \ldots, \infty\}$ then F is a dihedral group D_{2m}. Let $D = H_j^+ \cap H_k^+$. Then it is easy to see that the *sectors $f(D)$ for $f \in F$* tessellate V, as shown in the diagram below. We will denote the length of any f in the subgroup F by $l_F(f)$. (This is clearly no smaller than the length $l(f)$ in G.) Any chain of sectors from D to fD must have length at least $l_F(f)$.

$m = 5$

$$R_j R_k(D)$$
$$R_j(D)$$
$$D$$
$$R_k(D)$$
$$R_k R_j(D)$$
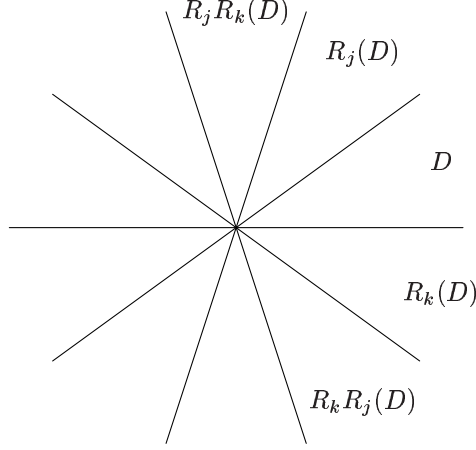
For g there is a minimal value of $l(fg)$ for $f \in F$, say $l(g_o)$ where $g = f_o g_o$ and $f_o \in F$. Note that $g' = R_k g$ so $l(g_o) \leqslant l(g') = L - 1$. Therefore we may apply Q_{L-1} to g_o. The choice of g_o ensures that

$$l(R_j g_o) \geqslant l(g_o) \text{ so } l(R_j g_o) = l(g_o) + 1 .$$

Then Q_{L-1} implies that $g_o C \subset H_j^+$. Similarly $g_o C \subset H_k^+$. So

$$g_o C \subset H_j^+ \cap H_k^+ = D$$

and consequently $gC \subset f_o D$. In particular this shows that gC is either a subset of H_k^+ or else of H_k^-. To complete the proof we will need the following lemma.

Lemma 6.2
With the notation introduced above we have

$$l(g) = l_F(f_o) + l(g_o) .$$

Proof:
It is clear that $l(g) \leqslant l(f_o) + l(g_o) \leqslant l_F(f_o) + l(g_o)$. We need to show that the reverse inequality holds. Consider a chain of cells

$$C = C_0, C_1, C_2, \ldots, C_L = gC \qquad (*)$$

with length $L = l(g)$ and C_{n+1} obtained from C_n by reflection in a side of C_n. Then there are elements $g_n \in G$ with $C_n = g_n C$ and $l(g_n) = n$. The hypothesis Q_{L-1} and our argument above show that each $g_n C$ lies in one of the sectors $(fD : f \in F)$. This gives a chain of sectors from D, which contains C, to $f_o D$, which contains gC. Of course several of the cells $g_n C$ may lie in the same sector fD, so the chain of sectors may have length less than L. However, its length, say d, must be at least $l_F(f_o)$.

If $g = g_o$, then $f_o = I$ and the lemma is clear. Otherwise, $l_F(f_o) \geqslant 1$ so the chain $(*)$ above must leave the sector D and pass into $R_j D$ or $R_k D$. Let us suppose, without loss of generality, that it passes into $H_j D$. So there is an $n < L$ with

$$C = C_0, \ C_1, \ C_2, \ \ldots, \ C_n \subset D \text{ but } C_{n+1} \subset R_j D .$$

This means that $C_{n+1} = R_j C_n$. Now consider the new sequence obtained by reflecting using R_j. This gives

$$C = C_0, \ C_1, \ C_2, \ \ldots, \ C_n = R_j C_{n+1}, \ R_j C_{n+2}, \ \ldots, \ R_j C_L = R_j gC$$

which is a chain of length $L - 1$ from C to $R_j gC$. We can repeat this process each time the chain $(*)$ passes from one sector to another and this happens d times with $d \geqslant l_F(f_o)$. This leads to a chain of length $L - d$ from C to $g_o C$. Hence,

$$l(g_o) \leqslant L - d \leqslant l(g) - l_F(f_o)$$

which proves that $l(g) \geqslant l_F(f_o) + l(g_o)$ as required. $\qquad \square$

Note that we can also apply the Lemma to the element $R_k g$. For we know that $gC \subset f_o D$, so $R_k g C \subset R_k f_o D$. Then the proof of the lemma shows that

$$R_k g = R_k f_o g_o \qquad \text{and} \qquad l(R_k g) = l_F(R_k f_o) + l(g_o) .$$

Now we can return to the argument for the Proposition. The diagram above makes it clear that the Proposition holds for F. Thus we have

$$f_o D \subset H_k^\pm \qquad \Leftrightarrow \qquad l_F(R_k f_o) = l_F(f_o) \pm 1 .$$

We already know that $gC \subset f_o D$ and, if part of a sector fD lies in the half-space $H_k^\pm$, then all of it must lie in that half-space. Hence

$$gC \subset H_k^\pm \qquad \Leftrightarrow \qquad f_o D \subset H_k^\pm .$$

Using the lemma and the remark immediately following it we see that

$$l_F(R_k f_o) = l_F(f_o) \pm 1 \qquad \Leftrightarrow \qquad l(R_k g) = l(g) \pm 1 .$$

Putting these three equivalences together we see that

$$gC \subset H_k^\pm \qquad \Leftrightarrow \qquad l(R_k g) = l(g) \pm 1$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We can now prove that the cells $g_1 C$ and $g_2 C$ are disjoint unless $g_1 = g_2$. Suppose that $g_1 \neq g_2$. Then $g = g_1^{-1} g_2 \neq I$. We can write $g = R_k g'$ for some k and some $g' \in G$ with $l(g) = l(R_k g') = l(g') + 1$. The Proposition now shows that $g'C \subset H_k^+$ and $gC \subset H_k^-$. However, $C \subset H_k^+$ so C and gC are disjoint. Consequently $g_1 C$ and $g_2 C = g_1(gC)$ are disjoint.